DOCUMENT RESUME

ED 473 610 TM 034 711

AUTHOR	McLean, James E., Ed.; Kaufman, Alan S., Ed.
TITLE	Research in the Schools.
INSTITUTION	Mid-South Educational Research Association, MS.; East Tennessee State Univ., Johnson City.
ISSN	ISSN-1085-5300
PUB DATE	2001-01-00
NOTE	187p.; The individual articles are available in the Current Index to Journals in Education (CIJE); see TM 524 852-867.
AVAILABLE FROM	Mid-South Educational Research Association, Warf-Pickle Hall, Room 418, Eastern Tennessee State University, Johnson City, TN 37614-1709.
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Research in the Schools; v8 n1-2 2001
EDRS PRICE	EDRS Price MF01/PC08 Plus Postage.
DESCRIPTORS	*Academic Achievement; Cooperative Learning; *Educational Research; Elementary Secondary Education; *Instructional Effectiveness; Preservice Teachers; *Research Methodology; Scholarly Journals; Statistical Significance; *Teacher Education

ABSTRACT

"Research in the Schools" publishes empirical studies focusing on the results of applied educational research, scholarly reviews of research, descriptions of technology applications and innovative teaching strategies, and other topics of interest to educational researchers. Issue number 1 contains these articles: (1) "Factors Associated with Reading Achievement and Attitude among Elementary School-Aged Students" (Pollyann J. Diamond and Anthony J. Onwuegbuzie; (2) "An Initial Investigation into the Effect of Decision-Making and Communication Practices on the Perceived Outcomes of Site-Based Management" (Scott C. Bauer); (3) "An Initial Reintegration Treatment of Children with Acute Lymphoblastic Leukemia (ALL)" (Michelle Lurie and Nadeen Kaufman); (4) "Correlates of School Academic Success: A State Report Card Study" (Jerry G. Mathews); (5) "An Empirical Study of Structural Relations between Science Education and Student Career Aspiration" (Jianjun Wang); (6) "The Role of Cooperative Learning in Research Methodology Courses: A Mixed-Methods Analysis" (Anthony J. Onwuegbuzie and Denise A. DaRos-Voseles); (7) "Likert Survey Primacy Effects in the Absence or Presence of Negatively-Worded Items" (J. Jackson Barnette); and (8) "Basic Cross-Validation: Using the 'Holdout' Method To Assess the Generalizability of Results" (Raquel M. Oxford and Larry G. Daniel). Issue 2 contains: (9) "The Relationship between Eighth-Grade Reading Scores and Achievement on the Georgia High School Graduation Test" (Deborah L. Demps and Anthony J. Onwuegbuzie); (10) "From Texas to Florida: Email Peer Coaching Dialogues among Preservice Teachers" (Suzanne I. Lapp); (11) "East Baton Rouge, School Desegregation, and White Flight" (Stephen J. Caldas and Carl L. Bankston, III); (12) "Science Achievement, Class Size, and Demographics: The Debate Continues" (Marie Miller-Whitehead); (13) "Characteristics of Effective Teachers: Perceptions of Preservice Teachers" (Ann E. Witcher, Anthony J. Onwuegbuzie, and Lynn C. Minor); (14) "Teachers; Beliefs about Mathematics Reform: Instructional Implications for Students with Learning Disabilities" (Kathleen M. T. Collins and Michael M. Gerber); (15) "Academic

Reproductions supplied by EDRS are the best that can be made from the original document.

Success as a Function of the Gender, Class, Age, Study Habits, and Employment of College Students" (William J. Lammers, Anthony J. Onwuegbuzie, and John R. Slate); and (16) "The Utility of Statistical Significance Testing in Psychological and Educational Research: A Review of Recent Literature and Proposed Alternatives" (Jeremy R. Sullivan). Each article contains references. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 473 610

RESEARCH IN THE SCHOOLS

A nationally refereed journal sponsored by the
Mid-South Educational Research Association
and East Tennessee State University.

Volume 8, Numbers 1 and 2

2001

TM034711

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

J.A. Boser

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

RESEARCH IN THE SCHOOLS

A nationally refereed journal sponsored by the
Mid-South Educational Research Association
and East Tennessee State University.

Volume 8, Number 1 **Spring 2001**

Factors Associated With Reading Achievement and Attitude Among Elementary School-Aged Students 1
Pollyann J. Diamond and Anthony J. Onwuegbuzie

An Initial Investigation into the Effect of Decision-Making and Communication Practices on the Perceived Outcomes of Site-Based Management 13
Scott C. Bauer

An Initial Reintegration Treatment of Children with Acute Lymphoblastic Leukemia (ALL) 29
Michelle Lurie and Nadeen Kaufman

Correlates of School Academic Success: A State Report Card Study 45
Jerry G. Mathews

An Empirical Study of Structural Relations Between Science Education and Student Career Aspiration 51
Jianjun Wang

The Role of Cooperative Learning in Research Methodology Courses: A Mixed-Methods Analysis 61
Anthony J. Onwuegbuzie and Denise A. DaRos-Voseles

Likert Survey Primacy Effect in the Absence or Presence of Negatively-Worded Items 77
J. Jackson Barnette

Basic Cross-Validation: Using the "Holdout" Method to Assess the Generalizability of Results 83
Raquel M. Oxford and Larry G. Daniel

James E. McLean and Alan S. Kaufman, Editors

RESEARCH IN THE SCHOOLS

Information for Authors

Statement of Purpose

RESEARCH IN THE SCHOOLS (ISSN 1085-5300) publishes original contributions in the following areas: 1) *Research in Practice*--empirical studies focusing on the results of applied educational research including cross-cultural studies, 2) *Topical Articles*--scholarly reviews of research, perspectives on the use of research findings, theoretical articles, and related articles, 3) *Methods and Techniques*--descriptions of technology applications in the classroom, descriptions of innovative teaching strategies in research/measurement/statistics, evaluations of teaching methods, and similar articles of interest to instructors of research-oriented courses, 4) *Assessment*--empirical studies of norm-referenced, criterion-referenced, and informal tests in the areas of cognitive ability, academic achievement, personality, vocational interests, neuropsychological functioning, and the like, and 5) *Other* topics of interest to educational researchers. *RESEARCH IN THE SCHOOLS* is devoted to research conducted in *any* educational setting from a conventional elementary school or high school to a training program conducted within an industry. Likewise, there are no age restrictions on the sample, since the educational settings may include preschools, continuing education classes for adults, or adaptive skills courses in nursing homes. Studies conducted in settings such as clinics, hospitals, or prisons are ordinarily inappropriate for *RESEARCH IN THE SCHOOLS* unless they involve an educational program within such a setting. One goal of *RESEARCH IN THE SCHOOLS* is to provide a training ground for graduate students to learn effective reviewing techniques. Consequently, the journal utilizes a Graduate Student Editorial Board composed mostly of students in educational psychology and educational research. Members of this Editorial Board, each sponsored by a professor, provide supplementary reviews for a selection of submitted articles, and receive both direct and indirect feedback of the quality of these reviews.

Preparing Manuscripts

Authors should prepare manuscripts in accordance with the stylistic rules and guidelines delineated in the *Publications Manual of the American Psychological Association* (4th ed., 1994), which is available from: Order Department, American Psychological Association, PO Box 2710, Hyattsville, MD 20784. Number the pages consecutively. All manuscripts will be subject to editing for sexist language.

Author Identification

Authors should put the complete title of the article on the first text page, but they should exclude their names. Subsequent pages should include only a running head. They should prepare a separate sheet with the complete title of the article and their names and affiliations; this procedure will ensure anonymity in the review process. Authors should supply addresses and phone numbers, and electronic mail addresses and fax numbers (if available), for potential use by the editorial staff and, later, by the production staff. Unless otherwise stated, the first-named author will be sent correspondence, galley proofs, copyright forms, and so forth.

Submission of Manuscripts

Submit manuscripts in triplicate to **James E. McLean, Co-Editor, RESEARCH IN THE SCHOOLS, Warf-Pickle Hall, Room 418, Eastern Tennessee State University, Box 70685, Johnson City, TN 37614-1709. Please direct questions to jmclean@etsu.edu**. All copies should be clear and readable; dot matrix is acceptable only if it meets these qualities of legibility. Length of the manuscripts, including references and tables, should ordinarily range from about 10 to 40 typed, double-spaced, 8-1/2 X 11-inch pages, using 11-12 point type. Abstracts are limited to 125 words. Brief reports of research are not encouraged. Authors are encouraged to keep a hard copy of the manuscript to guard against loss. It is assumed that all manuscripts submitted for publication are original material and have not been simultaneously submitted for publication elsewhere. When manuscripts are accepted for publication, authors are encouraged to submit the final version on a computer disk along with the hard copy.

Copyright and Permissions

Authors are granted permission to reproduce their own articles for personal use. Others must request permission to reproduce tables, figures, or more than 500 words of text from the editors. Copyright © 2000 by the Mid-South Educational Research Association.

Volume 8, Number 1 **Spring 2001**

EDITORS
James E. McLean, *East Tennessee State University*
Alan S. Kaufman, *Yale University School of Medicine*

ASSOCIATE EDITORS
Charles M. Achilles, *Eastern Michigan University*
James C. Kaufman, *Educational Testing Service*

PRODUCTION EDITOR
Margaret L. Rice, *The University of Alabama*

COPY EDITOR
Michele G. Jarrell, *University of Alabama at Birmingham*

EDITORIAL BOARD
Jan E. Alm, *Uppsala University (Sweden)*
Margery Arnold, *Trinity University*
Mark Baron, *University of South Dakota*
Alan W. Brue, *Carrollton City (Georgia) Schools*
Robin Cook, *Wichita State University*
Larry G. Daniel, *University of North Florida*
James M. Ernest, *State University of New York at Buffalo*
Daniel Fasko, Jr., *Morehead State University*
Ann Gallagher, *Educational Testing Service*
Glennelle Halpin, *Auburn University*
Harrison Kane, *University of Nevada, Las Vegas*
JinGyu Kim, *National Institute of Professional Administration (Korea)*
Jwa K. Kim, Middle *Tennessee State University*
Frank R. Lawrence, *Pennsylvania State University*
Robert E. Lockwood, *Department of Postsecondary Education (Alabama)*
Robert Marsh, *Chattanooga State Tech. Community College*
Jerry Mathews, *Auburn University*
Charles McLafferty, Jr., *University of Alabama at Birmingham*
Peter C. Melchers, *University of Cologne (Germany)*
Claire Meljac, *Hospital Sainte-Anne (France)*
Soo-Back Moon, *Catholic University of Hyosung (Korea)*
David T. Morse, *Mississippi State University*
Jack A. Naglieri, *Ohio State University*
Anthony J. Onwuegbuzie, *Valdosta State University*
Cecil R. Reynolds, *Texas A & M University*
Janet C. Richards, *The University of Southern Mississippi*
Michael D. Richardson, *Georgia Southern University*
John Kyle Roberts, *Baylor College of Medicine*
Dan Robinson, *University of Texas at Austin*
Christopher Skinner, *University of Tennessee*
John R. Slate, *University of Texas at El Paso*
Paula Stanley, *Radford University*
Bruce Thompson, *Texas A & M University*

GRADUATE STUDENT EDITORIAL BOARD
Keri L. Altig, *University of Nevada, Las Vegas*
Adam S. Bristol, *Yale University*
James M. Graham, *Texas A & M University*
Debbie Hahs, *The University of Alabama*
Keith Hyatt, *University of Nevada, Las Vegas*
Kevin M. Kieffer, *Texas A & M University*
Ginny Lane, *University of North Texas*
Lynn C. Minor, *Valdosta StateUniversity*
William T. Pepper, *Mississippi State University*
Catherine A. Stacy, *University of Texas at Austin*

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 1, 1-11

Factors Associated With Reading Achievement and Attitudes Among Elementary School-Aged Students

Pollyann J. Diamond and Anthony J. Onwuegbuzie
Valdosta State University

The purpose of the present study was to investigate reading achievement and attitudes as a function of grade, gender, ethnicity, and socioeconomic status (SES). Participants comprised 1,968 children who were enrolled in Kindergarten through fifth-grade schools in a school district in Georgia. With respect to achievement, an analysis of variance (ANOVA) revealed (1) a gender x ethnicity interaction ($\omega^2 = .05$); (2) a main effect for ethnicity ($\omega^2 = .24$); (3) a main effect for SES ($\omega^2 = .19$); and (4) a main effect for grade ($\omega^2 = .22$). Moreover, a quadratic trend was found for grade, with reading achievement declining consistently across Grades 2, 3, and 4, and increasing slightly between Grades 4 and 5. With regard to attitudes, an ANOVA revealed (1) an SES x ethnicity interaction ($\omega^2 = .05$); (2) a main effect for gender ($\omega^2 = .12$); and (3) a main effect for grades ($\omega^2 = .12$). A cubic trend emerged for grades, with reading attitudes increasing between Grades 1 and 2, decreasing between Grades 2 and 3, sharply declining between Grades 3 and 4, and slightly decreasing between Grades 4 and 5. The implications of these findings are discussed.

The National Research Council has identified reading as a central component to success in contemporary society (Snow, Burns, & Griffin, 1998). Indeed, reading has become increasingly important as the demand for higher levels of literacy continues to expand due to the transition from an industrial to an information society (Stedman & Kaestle, 1987). Thus, improving children's reading is a foremost challenge facing teachers. Unfortunately, many elementary school children have reading difficulties. Indeed, the gap between good readers and poor readers continues to be large. Thus, as noted by Snow et al. (1998), it is clear that knowledge of factors related to reading achievement is important. In response to this need, researchers have examined both cognitive and affective aspects of reading. Specifically, correlates of reading achievement and attitude have been examined. To date, the variables that have been related consistently to reading achievement are reading attitudes, gender, ethnicity, and socioeconomic status, whereas variables that have been related consistently to reading attitudes are gender, ethnicity, and grade level.

Reading Achievement and Attitudes

McKenna and Kear (1990) and McKenna, Kear, and Ellsworth (1995) found that attitudes toward reading were positively related to reading ability. In establishing norms for the Elementary Reading Attitude Survey among 18,185 students in Grades 1 through 6, McKenna and his colleagues (i.e., McKenna & Kear, 1990; McKenna et al., 1995) found that recreational and academic attitudes toward reading of high-achieving readers were statistically significantly higher than those of low-achieving readers. This finding supports earlier research by Lipsky (1983) and Roettger (1980), although it should be noted that Lipsky's research was limited by a small sample size ($n = 20$) and the fact that only fifth-grade boys were examined. Surprisingly, relatively few studies have examined the relationship between reading achievement and attitudes.

Gender and Reading Achievement

In a longitudinal study of reading achievement and gender differences, Flynn and Rahvar (1994) reported that low-achieving males and females occurred with relatively equal frequency among a sample of primarily White children. Achievement data were gathered in Grades 1 and 3. However, significantly fewer females were referred for learning problems in reading, suggesting a gender bias with regard to potentially learning-disabled females.

In an extensive review of the literature, Bond and Dykstra (1997) reported that the reading achievement of females was consistently higher than that of males, regardless of the age of the students, according to various research study results (i.e., Balow, 1963; Carroll, 1948; Gates, 1960; Heilman, 1961; Pauley, 1951; Tomplin,

---

Pollyann J. Diamond, Ed.S., NCSP, is a School Psychologist with the Valdosta City Schools and is currently a doctoral candidate at Valdosta State University (Georgia) in Curriculum and Instruction. Anthony J. Onwuebuzie is an Assistant Professor in the Department of Educational Leadership at Valdosta State University. Correspondence should be addressed to: Anthony Onwuegbuzie, Department of Educational Leadership, College of Education, Valdosta State University, Valdosta, Georgia, 31698. E:Mail: tonwuegb@valdosta.edu.

1957; and Waejen & Gramlis, 1963 as cited in Bond & Dykstra, 1997).

More recently, Donahue, Voelkl, Campbell, and Mazzeo (1999) documented that, at all grades, females have had significantly higher average reading scale scores than have males. Similarly, in an analysis of NAEP data collected since 1971, trends reported by Campbell, Voelkl, and Donahue (1999) revealed that female students have had higher reading achievement scores than had their male peers; moreover, this gender gap has remained relatively stable over time. Also, Coleman (1999), in the 1994 NAEP Trial Assessment report for Georgia, reported that the reading proficiency (achievement) for female nine-year-old public school students also was higher than that for their male counterparts.

Gender and Attitudes Toward Reading

Anderson, Tollefson, and Gilbert (1985) studied the attitudes toward reading of intellectually gifted students. All 276 students included in the investigation were high achievers, who were enrolled in Grades 1 through 12. Thirty percent of the sample was minority. Overall, females in the study reported statistically significantly more positive attitudes toward reading than did males. This gender difference was particularly true with regard to recreational reading.

In a study undertaken using a random sample of students enrolled in public schools in the Rocky Mountain region, Parker and Paradis (1986) found significant differences between males and females in Grades 4 through 6, with females reporting statistically significantly more positive attitudes toward reading than did males. The greatest change in attitude was noted in recreational or non-classroom reading. Specifically, females had statistically significantly more positive attitudes toward recreational reading than did males. These differences were particularly significant in Grades 4 and 5. Stevenson and Newman (1986) reported similar findings in an investigation using children who were followed from kindergarten through Grade 10 in Minneapolis, Minnesota schools. However, in this study, the difference in attitudes was not noted until the 10th grade, whereby females had significantly more positive attitudes toward reading than did males.

Shapiro (1980, 1990) noted that females began school with more positive attitudes toward reading than did males. In the earlier study (i.e., Shapiro, 1980), second-grade females taught by male teachers had the most positive attitudes toward reading, and male students with female teachers had the least positive attitudes toward reading. Overall, the attitudes toward reading of females were significantly more positive than were those of the male students. The second study (i.e., Shapiro, 1990) investigated gender differences in attitudes toward reading based on two instructional methods--basal reader and whole language approaches. Again, females were consistently found to have statistically significantly more positive attitudes toward reading than did males.

The works of Barnett and Irwin (1994) and Wallbrown, Levine, and Engin (1981) support the contention that attitudes toward reading of female students are more positive initially and with increased years of schooling than those of their male counterparts. Interestingly, the finding that females have more favorable attitudes toward reading than do males also has been documented in studies conducted outside of the United States. In particular, Davies and Brimber (1993), who conducted their research in a British primary school, reported that the attitudes of females aged 6-11 years toward reading and reading materials were more favorable than were those of their male counterparts. While the gender differences were significant at the younger ages, they did diminish as a function of age.

In a recent investigation using the Elementary Reading Attitude Survey (McKenna & Kear, 1990), McKenna et al. (1995) documented that females had significantly more positive attitudes toward recreational reading than did males, and that the gap between males and females widened with age. In the same study, females were found to have significantly and consistently more positive attitudes toward academic reading than did males, with attitudes of both groups declining at the same rate over time. In neither case could the difference be accounted for by differences in reading ability.

In a longitudinal study of primarily non-minority students, Kush and Watkins (1996) also identified a significant and consistent decline in attitudes toward reading across time. The attitudes of females were consistently more positive than were those of males, especially with regard to attitudes toward recreational reading. The researchers in this inquiry also used the Elementary Reading Attitude Survey to measure recreational, academic, and composite attitudes toward reading. These gender differences replicated those found by Danielson and Tighe (1994), who used the same measure of attitudes toward reading with at-risk elementary school-aged students. In general, studies consistently have supported the contention that attitudes toward reading are significantly more positive among females than among males.

Ethnicity and Reading Achievement

Using the 1994 NAEP Trial Assessment, Coleman (1999) reported that, in Georgia, the reading proficiency of White students was higher than that of African American and Hispanic students. No differences in reading proficiency were noted from 1992 to1994 between Hispanic and White students.

Subsequent studies conducted by the NAEP reported similar results. Specifically, Campbell et al. (1999)

reported that, although the racial gap narrowed between 1971 and 1996, White students in 1996 attained higher reading achievement scores than did their African American peers. Similarly, Donahue et al. (1999) reported that, in 1998, the average reading score for White 9-year-olds was statistically significantly higher than that attained by African American, Hispanic, and American Indian same-age peers.

Surprisingly, only one long-term study was located that examined the relationship between ethnicity and reading achievement. The National Assessment of Education Progress (NAEP) was mandated federally by Congress in the 1970s to assess the achievement of 9-, 13-, and 17-year-olds in reading and other academic areas, and to address trends over time. Valencia, Hiebert, and Kapinus (1992) reported that no change was present in the reading achievement of 9- and 13-year-olds between 1971 and 1990. However, when results were disaggregated by ethnicity, African American students at all age levels had significantly higher reading achievement in 1990 than in 1971, whereas the reading achievement of White students remained relatively stable over time. Overall, African American students were consistently lower in reading achievement than were White students.

Ethnicity and Attitudes Toward Reading

Only one study was identified that directly investigated ethnic differences in attitudes toward reading. In the cross-sectional norming study for the Elementary Reading Attitude Survey, McKenna et al. (1995) reported that the attitude toward recreational reading activities of all groups, Grades 1 through 6, became less positive with increased years of schooling, beginning as positive at Grade 1 and declining to relative indifference by Grade 6. Ethnicity appeared to be unrelated to the decline in attitudes toward reading. However, attitudes toward academic and recreational reading activities of African American students were significantly less positive than those of White students at every grade level. The attitudes of White students toward academic and recreational reading declined consistently with increased years of schooling, whereas those of African American students stabilized between Grades 5 and 6. Compared to the White children, the attitudes of African American students were less positive toward recreational reading, but more positive toward academic reading.

Socioeconomic Status and Reading Achievement

In a meta-analysis of research investigating the relationship between socioeconomic status and reading achievement, White (1982) reported differences in risk factors related to socioeconomic status. Specifically, when the individual student was used as the unit of analysis, socioeconomic status and reading achievement were only slightly related ($r = .23$ with verbal scores and $r = .27$ with composite achievement). However, when aggregate data were used as the unit of analysis, a strong relationship was noted between socioeconomic status and achievement ($r = .68$ with verbal scores and $r = .64$ with composite achievement).

Walberg and Tsai (1984) reported that 13-year-olds with higher socioeconomic status had higher reading achievement scores on the 1979-80 NAEP testing than did students from low socioeconomic status families. Included in the determination of socioeconomic status were factors such as educational level of parents and the presence of books and magazines in the home environment. Though these data are not current, findings are supported by more recent research using NAEP data (Campbell et al., 1999; Coleman, 1999; Donahue et al., 1999; Valencia et al., 1992). Additionally, Snow et al. (1998) suggested that students in schools where the majority of students have low socioeconomic status are at greater risk for poor reading achievement.

In an 18-year longitudinal investigation, Alwin and Thornton (1984) found a strong relationship between socioeconomic status and verbal ability, current placement, and amount of schooling that one receives. To a lesser effect, high school grade point average was found to be related to socioeconomic status. Effects in the early developmental years (i.e., preschool and early elementary school) were found to be stronger. The authors concluded that socioeconomic factors might have a greater impact on schooling outcomes, including reading achievement, at earlier ages.

In a more recent study, Walker, Greenwood, Hart, and Carta (1994) reported that children from low socioeconomic homes had lower scores on reading and spelling standardized achievement tests across the elementary grades. Differences were attributed to early language experiences that were associated with low socioeconomic status, with those from low socioeconomic homes being at a distinct disadvantage. In a similar, qualitative study, Purcell-Gates (1996) documented that the level of reading and writing in the home and community lives of families had a direct relationship with the student's understanding of written language as a system and, ultimately, reading achievement. However, this research was limited by the small sample ($n = 20$ families), use of a convenience volunteer sample, and non-randomization.

Pungello, Kupersmidt, Burchinal, and Patterson (1996) defined socioeconomic levels by lunch status, with free and reduced lunch status representing low socioeconomic level and full price status representing higher socioeconomic level. Results of the 4-year study involving 1,253 public elementary school students in a small southern city revealed deficits in reading achievement of both ethnic minority students and students of low socioeconomic status. Using a multiplicative risk factor model in which risk factors were treated individually and additively, low family income (i.e., free or reduced lunch status) and stressful events contributed to the prediction of low reading achievement test scores; African American students had significantly and consistently lower reading achievement test scores than did White students. Likewise, using a cumulative risk model in which stressful events were grouped and ranked in levels from 0 (least risk) to 3 (most risk) and correlated to reading achievement, similar results were noted. Specifically, students rated as having a higher cumulative risk score, combined with low socioeconomic status, obtained significantly lower reading achievement test scores than did their counterparts.

The results of the Pungello et al. (1996) study are consistent with the NAEP 1998 Reading Report Card for the Nation (Donahue et al., 1999). Using parental educational level to determine socioeconomic status, the average reading achievement score for students who reported higher parent educational levels was found by Donahue et al. (1999) to be significantly higher than for those who reported lower parental educational levels. These findings, in turn, are similar to the 1994 NAEP Trial Assessment (Coleman, 1999) report for Georgia.

Grade Level Effects on Attitudes and Achievement

In a cross-sectional investigation, McKenna et al. (1995) reported a significant decrease in attitudes toward reading among elementary school-aged students. These researchers found that, beginning in Grade 1, students' attitudes toward reading were relatively positive. By Grade 6, however, students' attitudes toward reading had become statistically significantly less positive—indeed, relative indifference toward reading appears to be common at this grade level. McKenna et al. reported large effect sizes pertaining to grade level of .54 for attitudes toward recreational reading and .80 for attitudes toward academic reading. Similar results were reported in a longitudinal investigation by Kush and Watkins (1996), who used the same attitude measure. Unfortunately, only one of these studies addressed changes in achievement over time.

Anderson et al. (1985), in a cross-sectional examination of high-achieving students in Grade 1 through Grade 12, found that students were significantly more positive toward reading at lower grade levels and became significantly less positive toward reading with increased years of schooling. Similarly, Shapiro (1990) documented that both males and females entered school with positive attitudes toward reading, but males' attitudes toward reading became significantly less positive with increased years of schooling. Thus, in general, studies consistently have supported the contention that attitudes toward reading become significantly less positive with increased years of schooling.

Summary

Some researchers in the area of reading achievement have used national data via the National Assessment of Educational Progress. However, these data are not collected to reflect progress in consecutive years, and are directed only to students in Grades 4, 8, and 12. Most of the other studies in this field, including those presented above, have tended to involve primarily non-minority populations from areas outside the Southeast region of the United States. Disturbingly, the sample sizes in many of these investigations have involved fewer than 200 students, with some researchers utilizing as few as 20 individuals. Thus, both the internal validity (i.e., arising from low statistical power) and external validity (i.e., the generalizability) of the findings of most of these investigations are questionable.

Because reading achievement and attitudes likely are reciprocally related, knowledge about factors that predict these outcomes would be informative. Yet, as noted by Kush and Watkins (1996, p. 315), "A more precise delineation of individual student characteristics and their interactive relationship with reading attitude and achievement remains to be discovered." This was the goal of the present study. Specifically, the purpose of the current inquiry was to investigate reading achievement as a function of grade, gender, ethnicity, and socioeconomic status (SES).

Additionally, we decided to examine predictors of reading attitudes. Specifically, their relationship to gender, ethnicity, socioeconomic status, and grade level were investigated. As such, this study replicates and extends existing research examining correlates of reading achievement and attitudes. Moreover, the current investigation is unique from the following two perspectives. First, a paucity of studies has examined simultaneously several potential correlates of reading achievement and attitudes. Second, although some researchers have utilized NAEP data to compare reading achievement at different points in time for selected grades (i.e., 4th, 8th, and 12th grades only), few or no investigators have examined changes in reading performance across consecutive grade levels. Thus, it was hoped that findings from this study would add to the current body of literature in the area of reading.

Methods

Participants

Participants comprised 1,968 children who were enrolled in four Kindergarten through third-grade schools and two fourth- and fifth-grade schools in a small inner city school district in Georgia. That is, all regular education students who were enrolled in these six schools in Grades 1 through 5, who participated in the March, 1998 administration of the Iowa Tests of Basic Skills, and who completed the Elementary Reading Attitude Survey in September, 1998, were included in the study. However, students who received services as students with intellectual disabilities were excluded from the sample because they did not consistently participate in the general curriculum and did not typically take the Iowa Tests of Basic Skills. Because of the small number of students (approximately 30) in the "Other" category, which included Hispanic, Native American, and Asian, these data were not included in the analysis. All demographic data were obtained from the 1997-98 Georgia Public Education Report Card (Georgia Department of Education, 1999). These data are reported in Table1.

Instruments

The Iowa Tests of Basic Skills is administered annually to all Kindergarten through eighth-grade students at the school system where the study took place. This instrument is a standardized test of academic achievement developed and published by The Riverside Publishing Company (The Riverside Publishing Company, 1994). Areas measured are Reading, Language, and Mathematics. Items are presented in a multiple-choice format. The score of interest in this study was the Reading Total. Normal Curve Equivalent scores, which were designed for comparisons across grades and test levels, were the scores used in the present study. The State of Georgia has adopted this particular test for use in all Georgia schools for state-mandated testing and has provided evidence that the instrument generates valid scores for use in the Georgia public schools (Georgia Department of Education, 1999). The Riverside Publishing Company reported reliability coefficients for the Reading Total, as measured by KR-20, for Grades 1 through 4 which ranged from .89 to .93.

The Elementary Reading Attitude Survey was developed by McKenna and Kear (1990). This survey yields three scores: attitudes toward recreational reading (Items 1-10), attitudes toward academic reading (Items 11-20), and full scale (summation of 20 items). The items on this Likert-type scale are weighted from 1 to 4, with a score of 4 representing the most positive response, and 1 representing the least positive response. Students choose one of four Garfield line drawings which are representative in expression from most to least favorable toward a given reading activity. Using norms which were initially established on a population of 18,185 students in 38 states in Grades 1-6, McKenna and Kear recommended that raw scores between 41 and 80 be indicative of increasingly positive attitudes toward reading, and scores 40 or less be considered increasingly negative attitudes toward reading.

McKenna and Kear (1990) reported reliability coefficients (i.e., Cronbach's alpha) of .80 or higher, except for the Recreational subscale in Grades 1 and 2 (i.e., .74 and .78, respectively). A factor analysis using the norm data supported the presence of two discrete scales reflecting different aspects of reading ability. Evidence of construct-related validity also was provided via positive correlations between the Recreational subscale and owning a library card and checking books out of the library, and a negative correlation between this subscale and time spent viewing television. Likewise, construct-related validity of the Academic subscale was evidenced through high scores on this subscale obtained by high ability readers. This survey tool was chosen by the Language Arts Curriculum Director to measure attitudes toward reading.

As recommended by many researchers (e.g., Onwuegbuzie, 1999; Onwuegbuzie & Daniel, 1999, 2000; Thompson & Vacha-Haase, 2000; Wilkinson & the APA Task Force on Statistical Inference, 1999), reliability coefficients always should be reported for the data at hand. Unfortunately, no reliability information was available for the Elementary Reading Attitude Survey for the current sample. Thus, instead, as recommended by Vacha-Haase, Kogan, and Thompson (2000), for each grade, the standard deviation of scores from the inducted study (i.e., McKenna & Kear, 1990) was compared to the standard deviation of scores for the present sample. Deviation scores, using the inducted study as the baseline, were as follows: Grade 1 (0.63); Grade 2 (0.71); Grade 3 (0.68); Grade 4 (1.12); and Grade 5 (0.04). Interestingly, these differences were all positive, suggesting that the current sample's reading scores were more variable than that of the inducted sample across all grade levels. However, with proportion deviations ranging from .01 to .10 ($M = 0.06$, $SD = 0.03$), it was concluded that the current sample was not too dissimilar from the inducted sample with respect to score variation on the Elementary Reading Attitude Survey. Indeed, using Magnusson's (1967) formula, which is based on the reliability of the inducted sample and the standard deviations of the inducted and present samples, the predicted reliability of the present sample's reading attitude scores was as follows: .88 for Grade 1, .89 for Grade 2, .89 for Grade 3, .91 for Grade 4, and .89 for Grade 5.

Procedure

The Iowa Tests of Basic Skills was administered in March 1998. Scores were available through a data management program called Student Data Management

System (1997). At the request of the Language Arts Curriculum Director, the Elementary Reading Attitude Survey was administered by classroom teachers during the week of September 21-25, 1998 to all first- through fifth-grade students in the six schools. These data were collated, summed into the two subscales and a total score, and matched with Iowa Tests of Basic Skills data for each student. This data-gathering tool also included a place for demographic information regarding ethnicity and gender. These data were coded with the attitudinal and achievement scores for each child. Lunch status was coded from information generated by the School Nutrition Program database. Once all data were collated, identifying information was removed.

Results

Table 1 presents the demographic distribution of the sample as a function of grade level. Descriptive statistics regarding attitudes toward reading and reading achievement for this sample are displayed in Table 2. In addition, the means and standard deviations of the reading achievement and reading attitude scores are presented in Tables 3-7, as a function of gender, ethnicity, socioeconomic status, and grade level, respectively.

Table 1
Demographic Distribution of Sample as a Function of Grade Level (Grades 1-5)

Grade	African American	White	Male	Female	Free/ Reduced	Full Pay	Total
1	75.8%	24.2%	51.6%	48.4%	78.4%	21.6%	
	316	101	215	202	327	90	417
2	77.6%	22.4%	48.2%	51.8%	82.5%	17.5%	
	354	102	220	236	376	80	456
3	78.3%	21.7%	47.3%	52.7%	78.8%	21.2%	
	321	89	194	216	333	87	410
4	76.6%	23.4%	48.7%	51.3%	68.7%	31.3%	
	269	82	171	180	241	110	351
5	76.3%	23.7%	52.1%	47.6%	68.9%	31.1%	
	255	79	174	159	230	104	334
Total	77%	23%	49.5%	50.5%	76.1%	23.9%	
	1515	453	975	993	1497	471	1968

Table 2
Descriptive Statistics for Students' Attitudes and Achievement Test Scores for the Total Sample

Scale	*M*	*SD*	*n*
Iowa Tests of Basic Skills:			
Reading Total	49.65	20.13	1550
Elementary Reading Attitude Survey:			
Attitude Total	63.02	12.03	1968

Table 3
Reading Achievement and Attitudes Toward Reading as a Function of Gender (Grades 2-5)

	Females		Males	
Scale	*M*	*SD*	*M*	*SD*
Reading Total	51.14	19.51	48.20	20.53
Attitude Total	65.59	11.02	60.26	12.41

Table 4
Reading Achievement and Attitudes Toward Reading as a Function of Ethnicity (Grades 2-5)

	African American		White	
Scale	*M*	*SD*	*M*	*SD*
Reading Total	44.93	17.88	65.72	19.07
Attitude Total	62.69	12.14	64.12	11.59

Table 5
Reading Achievement and Attitudes Toward Reading as a Function of Lunch Status (Grades 2-5)

	Free/Reduced		Full Price	
Scale	*M*	*SD*	*M*	*SD*
Reading Total	45.49	18.59	62.35	19.00
Attitude Total	62.85	12.03	63.32	12.03

Table 6
Reading Achievement Scores as a Function of Grade Level (Grades 2-5)

Grade	*M*	*SD*	*n*
2	58.06	19.39	455
3	49.99	20.91	410
4	43.18	16.82	351
5	44.57	19.29	334

Table 7
Attitudes Toward Reading Scale Scores by Grade Level

Grade	*M*	*SD*	*n*
1	65.15	12.03	417
2	65.22	12.11	456
3	64.00	11.58	410
4	59.97	12.12	351
5	59.34	10.84	334

Pearson's product-moment correlation coefficients were utilized to examine the relationship between reading achievement and attitudes toward reading among Grades 2-5 (Grade 1 students were not administered the same subtests of ITBS instrument and thus were excluded from all correlational analyses). Findings revealed a statistically significant relationship between these two constructs, r (1550) = .16, $p < .001$. Using Cohen's (1988) criteria, this relationship can be considered small. A follow-up series of correlations was examined for each grade (i.e., Grades 2-5). Using the Bonferroni adjustment for Type I error, the relationship between reading achievement and attitudes toward reading was not statistically significant for students in Grade 3 [r (410) = .01, $p > .05$] and Grade 5 [r (334) = .01, $p > .05$]. However, a statistically significant relationship between reading achievement and attitudes was found for Grade 2 [r (455) = .17, $p < .001$] and Grade 4 [r (351) = .17, $p < .001$]. Cohen's (1988) criteria suggest that these latter relationships are small.

A 2 (gender) x 2 (ethnicity) x 2 (socioeconomic status) x 4 (grade level) factorial analysis of variance (ANOVA) was conducted to examine reading achievement as a function of gender, ethnicity, socioeconomic status, and grade level. Findings revealed (1) a gender x ethnicity interaction [F (1, 1514) = 4.25, $p < .05$; ω^2 = .05]; (2) a main effect for ethnicity [F (1, 1514) = 90.34, $p < .001$; ω^2 = .24]; (3) a main effect for SES [F (1, 1514) = 53.66, $p < .001$; ω^2 = .19]; and (4) a main effect for grade [F (3, 1514) = 24.71, $p < .001$; ω^2 = .22]. A post-hoc Scheffé analysis revealed that, whereas African American females (M = 49.15, SD = 17.42) had statistically significantly higher reading scores (Cohen's d = 0.27) than did African American males (M = 44.36, SD = 18.19), no difference (Cohen's d = 0.07) emerged between White females (M = 60.28, SD = 18.88) and White males (M = 61.66, SD = 19.39).

Using Cohen's (1988) criteria, the effect sizes, as measured by ω^2, were moderate for the main effects and small for the interaction. Moreover, a quadratic trend was found for grade [F(1, 1578) = 22.01, $p < .001$], with reading achievement declining consistently across Grades 2, 3, and 4, and increasing slightly between Grades 4 and 5. The trend for reading achievement is shown in Figure 1.

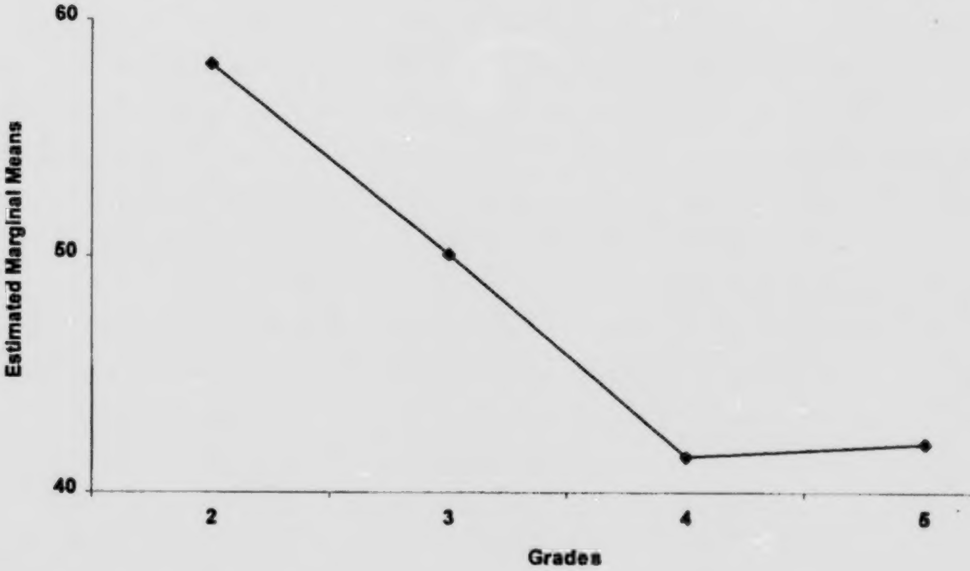


Figure 1. Estimated Marginal Means of Reading Total

A 2 (gender) x 2 (ethnicity) x 2 (socioeconomic status) x 5 (grade level) factorial analysis of variance (ANOVA) was conducted to examine attitudes toward reading as a function of gender, ethnicity, socioeconomic status, and grade level. Findings indicated (1) an SES x ethnicity interaction [*F* (1, 1924) = 5.75, $p < .05$; $\omega^2 = .05$]; (2) a main effect for gender [*F* (1, 1924) = 29.73, $p < .001$; $\omega^2 = .12$]; and (3) a main effect for grades [*F* (4, 1924) = 6.72, $p < .001$; $\omega^2 = .12$]. A post-hoc Scheffé analysis revealed that, whereas high-SES African American students (*M* = 64.07, *SD* = 12.83) had statistically significantly more positive reading attitudes (Cohen's *d* = 0.15) than did low-SES African American students (*M* = 62.22, *SD* = 12.08), no difference (Cohen's *d* = 0.22) emerged between high-SES White students (*M* = 63.28, *SD* = 11.65) and low-SES White students (*M* = 65.80, *SD* = 11.19).

Using Cohen's (1988) criteria, the effect sizes, as measured by ω^2, were in the small-to-moderate range. A cubic trend emerged for grades [*F* (1, 2007) = 8.20, $p < .001$], with reading attitudes increasing very slightly between Grades 1 and 2, decreasing between Grades 2 and 3, sharply declining between Grades 3 and 4, and slightly decreasing between Grades 4 and 5. The trend for reading attitudes is shown in Figure 2.

Discussion

The major purpose of the present study was to examine both reading achievement and attitude toward reading as a function of gender, ethnicity, socioeconomic status, and grade level among elementary school children. Also investigated was the relationship between reading achievement and attitudes toward reading. The use of a relatively large sample of students enrolled in a school system afforded the opportunity to add to the body of literature as related to that geographic area because of its high proportion of African American students and large number of students classified as low socioeconomic status.

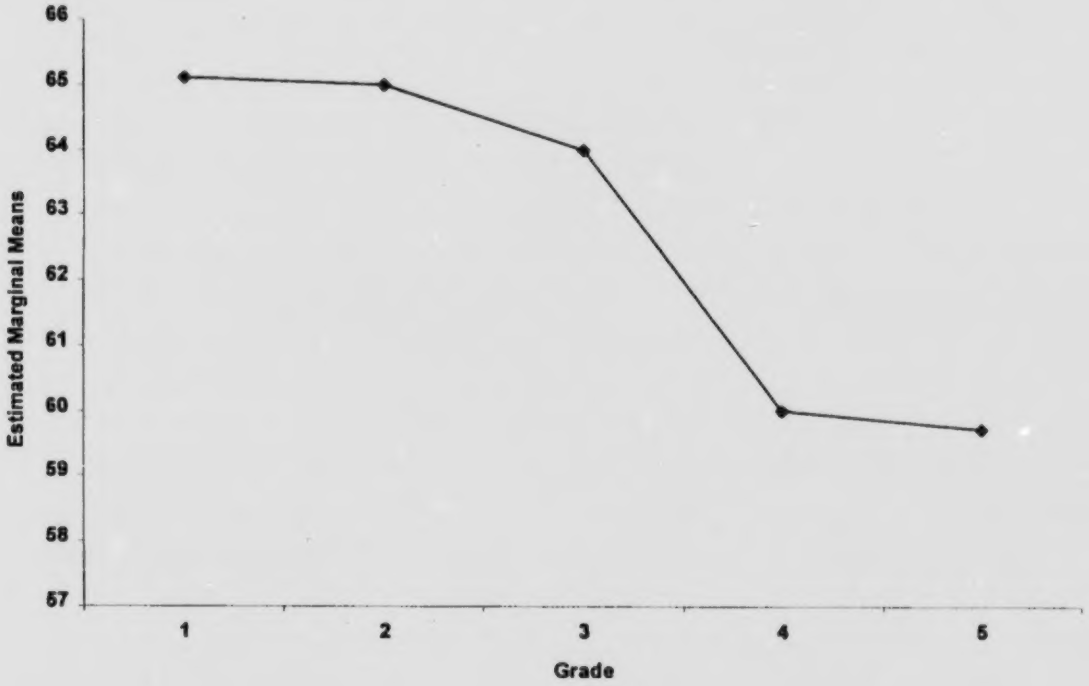


Figure 2. Estimated Marginal Means for Attitude Total

13

A small but statistically significant relationship was found between reading achievement and attitudes toward reading. Moreover, follow-up analyses suggested that this relationship was stronger for Grades 2 and 4 than for Grades 3 and 5. Future research should investigate the reliability of this finding. To the extent that this result is generalizable, it suggests that the effect of attitudes on reading achievement may be more important in some grades.

Ethnic differences were found with respect to reading achievement, with African American students exhibiting significantly lower levels of performance than did their White counterparts. The effect size pertaining to this difference was moderate. Indeed, this was true for students in Grades 2 through 5. This may suggest that African American students enter school with different literacy and reading experiences than do White children. Generally, these findings are consistent with the literature (McKenna et al., 1995).

Gender differences, in favor of females, were noted with respect to both reading achievement and attitudes. The effect size pertaining to the difference in reading achievement was moderate. This is consistent with the bulk of the literature (Anderson et al., 1985; Barnett & Irwin, 1994; Davies & Brimber, 1993; Parker & Paradis, 1986; Shapiro, 1980, 1990; Stevenson & Newman, 1986; Wallbrown et al., 1981). Interestingly, the interaction between gender and ethnicity revealed that, whereas no difference emerged between White males and White females, African American females outperformed their male counterparts. With respect to reading attitudes, the interaction found between SES and ethnicity suggests that the effect of SES on reading attitudes is a function of ethnicity. In the present study, reading attitudes discriminated African Americans on the low- and high-end of the SES continuum, but did not discriminate low-SES and high-SES White students. This finding suggests that the effect of SES on reading attitudes is more important for African American than for White students. In any case, future research should investigate further these interactions.

Students' attitudes toward reading were found to be significantly less positive beginning at the fourth grade than in Grades 1, 2, or 3. This finding is similar to what is reported in the literature (e.g., McKenna et al., 1995), with students becoming less positive about reading as they accrue experience with reading. The reason for this sudden difference is unknown but may be related to differing expectations for fourth grade students as compared with younger students. The trend found in reading attitudes likely explains the trend noted for reading achievement, since they were similar in nature. In any case, further investigation of these patterns is warranted.

Implications and Recommendations for Future Research

While of great practical application to the specific school system from which the data were generated, the results of this study should not be generalized to other school systems or geographic areas whose demographic variables differ significantly from those of the system from which the sample was selected. The sample was a nonrandom convenience sample, further limiting the generalizability of the results. However, the similarities noted to McKenna et al.'s (1995) results suggest that the sample in this study may not be too deviant from the national, stratified random sample in that study.

The three findings that appear to have the greatest theoretical and practical application relate to (a) the changes in attitude and achievement over years of schooling, (b) the aggregate effect of socioeconomic status on reading achievement, and (c) the gender x ethnicity interaction with respect to reading achievement. The major difference in achievement appears to occur between the end of second grade and the end of third grade (Grade 3 and Grade 4 for the school year in which the attitude survey was administered). Further investigation of this phenomenon may reveal other grouping structures or methodologies that would benefit the students and, possibly, intervene to reduce the decrease in scores and attitudes.

Likewise, the aggregate effect of socioeconomic status on reading achievement is important in a system where 76% of the students in Grades 1 through 5 are receiving free/reduced lunches. As noted above, all but one school has predominately free/reduced lunch status. This setting is the one identified in the literature as being detrimental to reading achievement because of the lack of appropriate role models available for students who seem to come to school with differing experiences and lower levels of literacy awareness (Snow et al., 1998). Because this has been reported in the literature and seems to have some effect within the system from which the sample was selected, the finding may address school attendance zone issues. The balanced literacy/balanced reading approach, which is the sole reading intervention currently being implemented at the school district under study, may improve children's literacy awareness so as to remove, at least in part, the detrimental effects of lower socioeconomic status that are reported in the literature.

Finally, the finding that African American females obtained higher levels of reading achievement than did their African American male counterparts is worthy of further investigation. Of particular interest is whether factors such as the lack of African American male elementary school teachers and the high proportion of single-parent, African American households without a constant father figure, play a role in stunting the reading achievement of African American boys. It is possible that these factors may result in African American boys receiving an inadequate level of encouragement or incentive to read either in the home or at school, which, in turn, may have dire consequences on their reading abilities. Stereotype threat, which is particularly common among African Americans, as evidenced by the view of many African American children that succeeding educationally at school is "acting white" (Kunjufu, 1988; Suskind, 1998), also has been found to debilitate cognitive test performance (Steele & Aronson, 1995). Thus, the potential debilitating role of stereotype threat on reading achievement among African American boys also should be examined. Thus, as noted by Onwuegbuzie (1997), much can be learned by examining further the within-ethnic differences found in the present investigation. Moreover, future studies are needed that include larger samples of other ethnic groups (e.g., Hispanic, Asian, and Native American).

Additional research also should consider qualitative data, such as interviews with the students, their parents, and the teachers involved in the implementation of the balanced literacy. Literacy leaders in each school would be a logical point of departure for collection of these data. Another important but often neglected area of research is analysis of data in middle grades, particularly with regard to content area reading skills and literacy in general. Indeed, recently, Demps and Onwuegbuzie (in press) found that eighth-grade reading scores on the Iowa Tests of Basic Skills (ITBS) statistically significantly predicted success on the five subtests (i.e., Writing, Language Arts, Math, Social Studies, and Science) of the Georgia High School Graduation Test (GHSGT). In this inquiry, eighth-grade reading scores predicted between 48% and 74% of the variance in scores on the subtests of the GHSGT. The proportions, which indicate a very large effect size (Cohen, 1988), suggest that success in reading also is crucial in the middle school years. Even if students learn to read in the lower grades, many of them do not make the transfer of skills to content area studies. As this is added to the continuum of reading research, follow-up investigations of students included in this original group also will be possible.

A limitation stems from the fact that these findings represent a snapshot of the reading achievement and attitudes toward reading of elementary students in a small inner-city school district. Although a valuable portrait in itself, additional color can be added through follow-up longitudinal studies. These investigations and replications in other geographic areas would add considerable knowledge to the literature about reading achievement and attitudes, which, in turn, may help educators to identify students who are most of need of interventions.

References

Alwin, D., & Thornton, A. (1984). Family origins and the schooling process: Early versus late influences of parental characteristics. *American Sociological Review, 49*, 784-802.

Anderson, M., Tollefson, N., & Gilbert, E. (1985). Giftedness and reading: A cross-sectional view of differences in reading attitudes and behaviors. *Gifted Child Quarterly, 29*, 186-189.

Barnett, J., & Irwin, L. (1994). The effects of classroom activities on elementary students' reading attitudes. *Reading Improvement, 31*, 113-121.

Bond, G., & Dykstra, R. (1997). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 32*, 348-427.

Campbell, J., Voelkl, K., & Donahue, P. (1999, February 28). Report in brief: NAEP 1996 trends in academic progress. [WWW document] URL http://nces.ed.gov/nationsreportcard/96report/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: John Wiley.

Coleman, B. (1999, February 28). The 1994 NAEP Trial Assessment [WWW document] URL http://nces.ed.gov/nationsreportcard/y25sa/tsa94.shtml

Danielson, K., & Tighe, P. (1994). Generating response to literature with at-risk third grade students. *Reading Horizons, 34*, 257-278.

Davies, J., & Brimber, I. (1993). Comics or stories? Differences in the reading attitudes and habits of girls and boys in years 2, 4, and 6. *Gender and Education, 5*(3), 305-320.

Demps, D.L., & Onwuegbuzie, A.J. (in press). The relationship between eighth-grade reading scores and achievement on the Georgia High School Graduation Test. *Research in the Schools*.

Donahue, P., Voelkl, K., Campbell, J., & Mazzeo, J. (1999). *NAEP 1998 reading report card for the nation and the states*. Washington, D. C.: U.S. Department of Education Office of Educational Research and Improvement.

Flynn, J., & Rahvar, M. (1994). Prevalence of reading failure in boys compared with girls. *Psychology in the Schools, 31*, 66-71.

Georgia Department of Education (1999, January 21). *1997-98 Georgia public education report card.* [WWW document] URL http://accountability.doc.k12.ga.us/4d/RCP/GaPage1.gif.

Kunjufu, J. (1988). *To be popular or smart: The black peer group*. Chicago, IL: African American Images.

Kush, J., & Watkins, M. (1996). Long-term stability of children's attitudes toward reading. *Journal of Educational Research, 89*, 315-319.

Lipsky, J. (1983). A picture-story technique to uncover covert attitudes associated with reading failure. *Reading Psychology, 4,* 151-155.

Magnusson, D. (1967). *Test theory.* Boston, MA: Addison-Wesley.

McKenna, M., & Kear, D. (1990). Measuring attitude toward reading: A new tool for teachers. *The Reading Teacher, 43,* 626-639.

McKenna, M., Kear, D., & Ellsworth, R. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly, 30,* 934-956.

Onwuegbuzie, A.J. (1997, April). *The impact of magnet schools on within-race academic achievement.* Paper presented at the annual meeting of the National Association of School Psychologists, Anaheim, CA.

Onwuegbuzie, A.J. (1999, September). *Common analytical and interpretational errors in educational research.* Paper presented at the annual meeting of the European Conference on Educational Research, Lahti, Finland.

Onwuegbuzie, A.J., & Daniel, L.G. (1999, November). *Uses and misuses of the correlation coefficient.* Paper presented at the annual conference of the Mid-South Educational Research Association, Point Clear, AL.

Onwuegbuzie, A.J., & Daniel, L.G. (2000, April). *Common analytical and interpretational errors in educational research.* Paper presented at the annual conference of the American Educational Research Association, New Orleans.

Parker, A., & Paradis, E. (1986). Attitude development toward reading in grades one through six. *Journal of Educational Research, 79,* 313-315.

Pungello, E., Kupersmidt, J., Burchinal, M., & Patterson, C. (1996). Environmental risk factors and children's achievement from middle childhood to early adolescence. *Developmental Psychology, 32,* 755-767.

Purcell-Gates, V. (1996). Stories, coupons, and the TV Guide: Relationships between home literacy experiences and emergent literacy knowledge. *Reading Research Quarterly, 31,* 406-428.

Roettger, D. (1980). Elementary students' attitudes toward reading. *Reading Teacher, 33,* 451-453.

Shapiro, J. (1980). Primary children's attitudes toward reading in male and female teachers' classrooms: An exploratory study. *Journal of Reading Behavior, 12,* 255-257.

Shapiro, J. (1990). Sex-role appropriateness of reading and reading instruction. *Reading Psychology, 11,* 241-269.

Snow, C., Burns, S., & Griffin, P (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, D.C.: National Academy Press.

Stedman, L., & Kaestle, C. (1987). Literacy and reading performance in the United States from1880 to the present. *Reading Research Quarterly, 22,* 8-46.

Steele, C.M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.

Stevenson, J., & Newman, R. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57,* 646-659.

Student Data Management System [Computer software]. (1997). Chicago, IL: The Riverside Publishing Company.

Suskind, R. (1998). *A hope in the unseen.* New York: Broadway books.

The Riverside Publishing Company (1994). *Technical summary I: Riverside 2000.* Chicago, IL: The Riverside Publishing Company.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60,* 174-195.

Vacha-Haase, T., Kogan, L.R., & Thompson, B. (2000, April). *Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Valencia, S., Hiebert, E., & Kapinus, B. (1992). National Assessment of Educational Progress: What do we know and what lies ahead? *The Reading Teacher, 45,* 730-734.

Walberg, H., & Tsai, S. (1984). Reading achievement and diminishing returns to time. *Journal of Educational Psychology, 76,* 442-451.

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development, 65,* 606-621.

Wallbrown, F., Levine, M., & Engin, A. (1981). Sex differences in reading attitudes. *Reading Improvement, 18,* 226-234.

White K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91,* 461-481.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 1, 13-27

An Initial Investigation into the Effect of Decision-Making and Communication Practices on the Perceived Outcomes of Site-Based Management

Scott C. Bauer
University of New Orleans

This study explores the relationship between the decision-making and communication practices of site-based teams and the perceived outcomes of site-based management. Survey data (N=133) are used to develop measures relating to the decision-making and communication processes used by participants engaged in site-based management and the resources available to site-based teams. Results of blocked regression analyses show that measures relating to the decision-making and communication practices have a strong impact on stakeholders' assessment of the effectiveness of site-based management at enhancing stakeholder influence, improving decision-making, and promoting educational outcomes, and that this impact holds when controlling for those factors most often cited in the literature as predictive of site-based management outcomes, i.e., factors relating to the resources provided to site teams. Findings suggest that it is important to account for decision-making and communication practices when conducting research into the efficacy of site-based management.

Although there is considerable support for the concept of site-based management, there is skepticism regarding whether restructuring decision-making can fulfill the promise of promoting school improvement (Malen & Ogawa, 1992; Ogawa & White, 1994). Studies show that there is seldom an explicit connection between the implementation of site-based decision-making and school performance (Cohen, 1988; Murphy & Beck, 1995; Taylor & Bogotch, 1994). Likewise, studies fail to show consistent support for the connection between the implementation of site-based management and intermediate benefits such as improved staff morale, stakeholder influence, and the use of quality planning practices (see, for example, David, 1989; Lindquist & Mauriel, 1989; Malen, Ogawa, & Kranz, 1990b; Murphy & Beck, 1995).

The overwhelming consensus in the literature on existing site-based management programs is that districts and schools seldom fully implement site-based systems (Marsh, 1994; Wohlstetter & Odden, 1992). Issues of "insufficient capacity" are often cited as explaining the failure of site-based management. "Capacity" equates to district support for site teams in terms of providing authority, training, time, information and other resources necessary to team operation. Districts rush to implement site-based management without considering what it takes to make the transition from traditional decision-making structures (Glickman, 1990).

The literature on site-based management is itself deficient in many regards. First, even as researchers assert that site-based management is poorly defined and that there is no single, best approach to implementing the process, a single model tends to be stressed in the literature, defining site-based management as devolving authority over issues relating to budget, staffing, and certain aspects of curriculum to a council made up of the building principal, teachers, other school staff, and parents (Bauer, 1998). Second, there are few systematic studies of the implementation of collaborative decision-making processes. Much of the literature consists of advocacy pieces associated with a district's implementation of the process, plan descriptions, and anecdotal accounts of what works (Malen, Ogawa, & Kranz, 1990a). The ambiguous nature of the subject and the fact that sites define site-based management differently make it hard to compare those studies that exist. The commission reports advocating adoption of decentralized decision-making offer little suggestion as to the steps needed to implement it (Conley & Bacharach, 1987).

Site-based management is justified as a reform in terms of two theoretical propositions (Malen et al., 1990b). Briefly, the first deals with motivation of stakeholders and holds that involvement in decision-making results in participants being more committed and enthusiastic about shared decisions than about decisions that are handed down to them hierarchically. The second relates to information processing and holds that those closest to students have better and more timely information available to them about what students need, thus devolving

---

Scott C. Bauer (Ph.D., Cornell University) is an assistant professor in the Educational Leadership, Counseling, and Foundations Department in the College of Education at the University of New Orleans, and Director of Research at the School Leadership Center of Greater New Orleans. Correspondence regarding this article should be directed to Scott C. Bauer, Educational Leadership, Counseling, and Foundations, 348 Bicentennial Education Bldg., University of New Orleans, New Orleans, LA 70148 or via e-mail to sbauer@uno.edu.

decisions closer to the point of delivery of educational services should result in better decisions and more timely delivery of needed services (Bauer, 1998; Murphy & Beck, 1995; Shedd & Bacharach, 1991).

Both of these theoretical propositions involve a complex set of interconnected events which are purported to occur between the initiation of stakeholder involvement and an eventual impact on educational outcomes: Participants must engage in dialogue and feel more involved, data must be shared, goals must be developed, those individuals most important to implementation must be involved, others must be kept informed, teams must be provided with needed resources such as time and information, and so on. These linkages are spelled out in considerable detail in theory (see, for instance, Murphy & Beck, 1995), yet involvement processes and the social-psychological effects purported to occur along the way are treated as a black box (Sharpe, 1996). That is, the complex decision-making and communication processes that occur between involvement and outcome are seldom explicitly measured or accounted for in empirical work on site-based management. Literature on site-based management seldom addresses the implementation process itself (Cotton, 1993; Miles & Louis, 1990), focusing instead on reviewing extant programs in terms of their progress in meeting stated goals. There are few empirical attempts to model the connections between these elements; studies seldom thoroughly describe exactly what site teams do, how they are configured, and how teams operate. The actual process of planning and implementing site-based management is seldom addressed.

To fully understand the nature of the connection between stakeholder involvement and school improvement, it is essential to develop ways to operationalize the sequence of events which are theorized to occur as administrators, staff and parents engage in the practice of site-based management. Further, to adequately assess the efficacy of site-based management as a reform, it is important to develop measures relating to the practice of site-based management and test whether there is empirical support for the theoretical propositions that connect them.

The research presented in this paper represents an initial attempt to respond to this gap in the literature. Specifically, survey data collected as a part of the evaluation of one city's three-year site-based management pilot project is used to develop measures of key elements relating to the decision-making and communication processes used by participants engaged in site-based management and the resources available to site-based teams. A model is then tested to determine the impact of these factors on the perceived effectiveness of site-based management and respondent's satisfaction with aspects of the pilot project.

Measuring Site Team Practices: Conceptual Framework

Although few researchers have attempted to categorize or describe dimensions relating to the actual operation of site-based management teams, Shedd (1987) and Shedd and Bacharach (1991) proposed the notion that collaborative decision-making processes can be described in terms of four broad dimensions:

- *Scope* refers to the nature of the subjects that site council participants discuss and includes such factors as whether agreed-upon goals exist, what types of issues councils address and how much power they have over these issues, and the limits on site team authority;
- *Formal structure* deals with which stakeholders are involved on site teams and how their roles are distributed. Structure includes such considerations as how many people serve on site teams, the mix of stakeholders, how individuals are selected to be on teams, and how the process is coordinated among schools and groups;
- *Decision-making process* deals with how the site teams actually go about making decisions and includes considerations like how they arrive at final decisions and the methods teams use for conflict resolution; and
- *Support* includes many of the issues referred to in the literature as "capacity," including whether adequate information, time and training are provided, whether management supports the site-based project, and whether teams have adequate authority to make decisions.

These dimensions were an attempt at describing the design of site-based management and the practices engaged in by site councils, but the model was not empirically tested.

The present study builds on this work. Survey items were constructed to tap participants' perceptions of items relating to scope, structure, process, and support in order to develop measures of site team practice. The conceptual framework was slightly modified, however. Shedd's (1987) original framework treated scope, structure, process, and support as four equally important dimensions together describing the design and practice of site-based management. The literature suggests, though, that dimensions dealing with the *practice of* site-based management and those that deal with the *resources provided to* site-based councils may differ in terms of their impact on council efficacy (see, for example, Bauer, 1998; Murphy & Beck, 1995). Variables related to support can be considered separately from the internal workings of the councils. The former deals with resources *provided to* councils, while the latter deal directly with the *operation of* school councils.

Research Questions

Two primary research questions were addressed in this study. First, what is the relationship between measures relating to the decision-making and communication practices of site-based teams and the perceived outcomes of site-based management? Since these practices are so seldom dealt with explicitly in research on site-based management, I was interested to determine whether measures relating to communication and decision-making practices would have a statistically significant relationship with various outcomes, and whether different factors related to practice would have an impact on different outcomes. Second, what is the relationship between factors relating to site council practices and perceived outcomes when controlling for those factors most often cited in the literature as predictive of site-based management outcomes (i.e., factors relating to the resources provided to site teams)? The second question asks about the relative importance of site council practices and resources to perceived outcomes.

Model and Hypotheses

Hypotheses regarding the relationship between communication and decision-making practices engaged in by stakeholders and the perceived outcomes of site-based management are somewhat difficult to formulate since these are so seldom mentioned explicitly in research on site-based management. The literature more broadly focused on the use of teams in work organizations, however, often deals with these factors [see Guzzo & Dickson (1996) or Bettenhausen (1991) for reviews]. For example, the component ideas embodied in scope include such things as goal consensus, agreement on the types of issues councils address, and the amount of authority they have over these issues; structure includes issues of coordination; and process involves interpersonal dynamics of teams. Parker (1994) isolated issues like goal ambiguity, interpersonal dynamics, and boundary management or coordination as important to the success of cross-functional teams in industry. He also discussed ambiguity relating to the scope of decision authority. In a separate study, Parker (1990) emphasized issues relating to communication and listening skills, conflict resolution procedures, and decision-making processes. In their volume on exemplary teams, Katzenback and Smith (1993) observed that commitment to a common purpose and performance goals is critical to team success, as is the development of a common approach to accomplishing their work. Adams (1988) discussed the problems associated with "unaligned teams," by which he means teams that have not agreed on purpose and the team's mission, and in their study of why teams fail, Robbins and Finley (1995) also highlighted issues relating to agreement on purpose. Douglass and Douglass (1992) discussed issues relating to team process and trust, issues that also played prominently in Douglas McGregor's description of effective teams nearly thirty years ago. McGregor defined trust as knowing that another person would not take unfair advantage of you, and he connected trust to issues of mutual support and open communications among team members (Burke, 1988). Thus, the literature on the use of teams in work settings provides a reasonable foundation for the expectation that scope, structure and process will be positively related to outcomes like the perceived effectiveness of site-based management and council member satisfaction.

> H_1: There will be a statistically significant, positive relationship between factors relating to the communication and decision-making practices engaged in by site-based councils and perceived outcomes of site-based management.

The literature on site-based management cited earlier provides a solid basis for formulating hypotheses regarding the relationship between measures of the resources provided to site-based councils and outcomes. Issues relating to the provision of adequate time, training, and authority are frequently mentioned as explanations regarding the failure of site-based management in many sites, and issues relating to administrative support are also widely mentioned as critical to the success of site-based processes (Bauer, 1998; Murphy & Beck, 1995). Hence, the following relationship is expected between measures relating to the resources provided to site-based councils and perceived outcomes:

> H_2: There will be a statistically significant, positive relationship between factors relating to the resources provided to site-based councils and the perceived outcomes of site-based management.

In fact, based solely on a reading of the literature, we might expect that these issues would emerge as dominant predictors of various outcomes.

Two additional variables were included in the analysis to account for the effects of two issues that are mentioned in the literature on site-based management. First, a dummy variable was entered which measures the "newness" of the site council. As described in the next section, members of twelve site-based councils are included in this study, and of these four of the councils were operating for approximately one year when they were surveyed while the remaining teams were in their third year of operation. On the one hand, Kanter (1983) suggests that especially for employees who are not accustomed to involvement, the initiation of a participation process can act as a sort of elixir, renewing and rejuvenating those involved. On the other hand, the literature suggests that site-based teams typically experience a period of confusion (Kirby, 1992; Murphy & Beck, 1995), which would suggest that newness would be a liability. Relying on this literature, the following is expected:

> H_3: There will be a statistically significant, negative relationship between the newness of a site-based management team and the perceived outcomes of site-based management.

A second control variable was constructed to control for the school type. The literature on site-based management suggests that it is more difficult to implement the process in larger and more administratively complex schools. For instance, Akin (1992) observed that the traditional high school structure contains several elements that stifle the development of a positive culture, including large staffs, isolation, departmentalization of staff and tracking of students. Hatry, Morley, Ashford, and Wyatt (1993) remarked on the fact that the departmentalization and size of high schools creates issues that negatively impact on the establishment of site-based management, and in a study of New Jersey school districts, Corcoran, Hansen, and Shidlowski (1988) found support for the notion that organizational complexity reduces the likelihood of involvement. Thus, it is reasonable to predict:

> H_4: There will be a statistically significant, negative relationship between administrative complexity and the perceived outcomes of site-based management.

The hypotheses are summarized in graphic form in Figure 1.

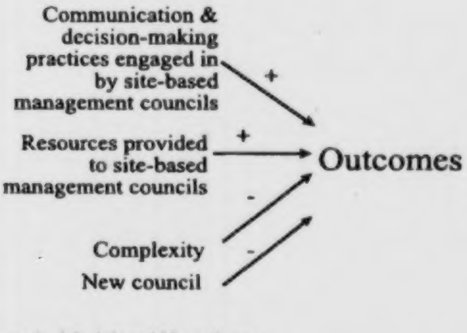

Communication & decision-making practices engaged in by site-based management councils
+
Resources provided to site-based management councils
+
Outcomes
-
Complexity
-
New council

Figure 1. Model and Hypotheses.

Participants

The data presented were collected as a part of an interim evaluation of the site-based management pilot project in Metro (a pseudonym), a major Midwestern city that had implemented its site-based pilot project three years prior to the evaluation. Initially, eight of Metro's schools were involved in the pilot project. One year prior to the study, four additional pilot schools were added, for a total of twelve schools in the sample. Surveys were distributed through a district steering committee charged with evaluating the pilot program. Each respondent was given a cover letter describing the project, a survey form, and a response envelope designed to ensure confidentiality. Surveys were collected by members of the district committee.

All members of the school-site councils were surveyed. Of the twelve school-site councils, four were secondary schools and eight were elementary schools. A usable sample of 133 surveys were returned from a total of 208 distributed, for a response rate of 64%. From this total, 87 surveys (65%) were from elementary site council members and 46 were from secondary site council members (35%). An examination of district records revealed that the responses from among the various stakeholder groups represented on site councils reflects the composition of the councils (i.e., no stakeholder group is underrepresented in the present sample).

It is important to note that Metro took a very typical approach to implementing site-based management, at least in the sense that the central administration determined that it would be beneficial to establish a pilot project involving decentralized decision-making and provided very little further guidance to the schools. Central administrators selected pilot schools, provided some rudimentary training in group process skills for key stakeholders in these schools, but otherwise left it up to each pilot school to design its own site-based processes. The assumption was that each school could fill in the details and establish their own rules of practice, and that it would be an infringement on school autonomy to specify how each school should or could operate. At the time, several of these councils felt stalled; a high level of distrust existed and the central office's motives in supporting site-based management were questioned.

Analysis

In order to address the research questions, survey items were developed to measure factors relating to communication and decision-making practices and the resources provided to site-based councils. The survey was pilot tested with school councils from a different school system prior to implementation in Metro. Survey items were subjected to factor analysis using principal components extraction with varimax rotation. Factors with an eigenvalue over 1.0 were extracted. The rotated factor loadings are presented in the next section.

To examine the relationships between site council practices, resources, and perceived outcomes, blocked regression models were constructed. To determine the relationship between communication and decision-making practices and perceived outcomes, the factors relating to these measures were entered first. To test whether measures relating to site council communication and decision-making practice have a statistically significant impact on outcomes when accounting for the effects of the resources provided to site teams, factors relating to resources were entered next, along with the two control variables measuring administrative complexity and whether the respondents served on a new council. The change in R^2 is examined as a measure of the contribution of each block in accounting for the variance in the outcome measures. Unstandardized and standardized regression parameters are presented, and results of the statistical significance test (t test) are reported using an alpha level of .05. In addition, partial eta^2 was computed for each univariate t test to measure effect size. Stevens (1996) notes that partial eta^2 is a consistent measure of effect size that is applicable to F and t tests, and that for samples of about 50 or more, partial eta^2 and eta^2 differ very little. His recommendation for interpreting the effect size was also adopted: .01 is treated as a small effect, .06 a medium effect, and .14 a large effect.

Finally, analysis revealed that using listwise deletion of cases with missing values in the regression analysis resulted in the loss of up to 25 cases, or nearly 20% of the total amount of data. Examination of the data showed, however, that no individual item had more than four missing values. As Johanson, Green and Williams (1999) observed, in cases when data contain a small number of missing values, it is reasonable to deal with this problem by imputing means. Given the small number of missing values for individual items, balanced against the large number of cases that would be lost using listwise deletion, scales used in the regression analysis were constructed using items that had means substituted for missing values.

Instruments

Survey items were constructed to measure the dimensions represented in the conceptual framework, namely scope, structure, process, and support, along with measures relating to the outcomes associated with site-based management. Consistent with this framework, factors dealing with the decision-making and communication practices engaged in by councils (scope, structure, and process) are discussed separately from those relating to the resources provided to site-based management teams (support). Items dealing with outcomes are then presented, and finally, the two control variables are described.

Although the scales created for this study are new, and the factor analysis is exploratory in nature, any survey is designed based on certain assumptions about the underlying concepts being measured. In presenting these items, our assumptions about the underlying factor pattern based on existing theory and research are presented as a heuristic, followed by the results of the factor analysis and a description of the scales created for the study.

Decision-Making and Communication Practices

Items relating to the decision-making and communication practices engaged in by site councils have to do with the internal workings of site councils, which in turn relate to the design of site-based management and the processes used by stakeholders. To tap these issues, Metro respondents were asked to rate the accuracy of each of the following questions dealing with the operation of their school site council (1=very inaccurate, 2=somewhat inaccurate, 3=somewhat accurate, 4=very accurate).

a. Site council members and those in authority agree on what kinds of decisions the council may and may not make.
b. Site council members have a clear sense of the goals they want to achieve.
c. The site council makes effective use of research bearing on issues it addresses.
d. The site council is creative in how it addresses issues.
e. The site council has real influence on issues of importance.
f. Site council activities and those of other committees are well coordinated.
g. All members of the site council have an equal opportunity to be involved in decisions.
h. The site council keeps those who might be affected by decisions informed of its progress.
i. The site council gives those who might be affected by decisions opportunities to have input.
j. The site council membership is representative of the staff in the building.
k. Members of site council listen to each other and are prepared to change their opinions.
l. Members of the site council trust one another.
m. Site council members communicate openly and honestly during meetings.
n. Even when members disagree, they focus on what they believe to be best for students.
o. Members support site council decisions outside the council meetings.
p. The site council takes responsibility for its decisions.

Based on the conceptual framework presented earlier, three factors were expected to emerge concerning site council decision-making and communication practices: items a - e relate to scope, items f - j represent structure, and items k - p involve decision processes.

Table 1 shows that three factors emerged from the analysis. Item c, which deals with whether members feel that their council makes effective use of research; item f, which measures whether respondents feel that site council activities and those of other committees are well coordinated; and item p, which deals with whether respondents feel that site council members take responsibility, load on more than one factor, and are omitted from the analysis.

Table 1
Factor Analysis of Communication and Decision-Making Practice Items

		1	2	3
a.	Site council members and those in authority agree on what kinds of decisions the council may and may not make.	**.810**	-.002	.184
b.	Site council members have a clear sense of the goals they want to achieve.	**.683**	.279	.373
c.	The site council makes effective use of research bearing on issues it addresses.	.535	.429	.327
d.	The site council is creative in how it addresses issues.	**.659**	.183	.474
e.	The site council has real influence on issues of importance.	**.827**	.235	.222
f.	Site council activities and those of other committees are well coordinated.	.558	.422	.435
g.	All members of the site council have an equal opportunity to be involved in decisions.	.284	**.523**	.393
h.	The site council keeps those who might be affected by decisions informed of its progress.	.404	**.583**	.458
i.	The site council gives those who might be affected by decisions opportunities to have input.	.442	**.629**	.319
j.	The site council membership is representative of the staff in the building.	.023	**.856**	.082
k.	Members of the site council listen to each other and are prepared to change their opinions.	.338	.219	**.818**
l.	Members of the site council trust one another.	.401	.198	**.782**
m.	Site council members communicate openly and honestly during meetings.	.217	.166	**.779**
n.	Even when members disagree, they focus on what they believe to be best for students.	.396	.192	**.768**
o.	Members support site council decisions outside the council meetings.	.219	.345	**.666**
p.	The site council takes responsibility for its decisions.	.121	.511	.577

The first factor deals with scope, and it includes issues relating to the power council members believe they have, their sense of council members' agreement on goals, and their sense of whether council members and those in authority agree on the decisions site councils may and may not make. The item dealing with whether council members feel they are creative in their decision-making also emerges in this factor, tapping into perceptions of the quality of decision making. A four-item scale, including items a, b, d, and e from Table 1, was created for this factor, which is labeled *scope*. Cronbach's alpha for the scale is .86.

The second factor relates to formal structure. The items that load on this factor deal with coordination, both in terms of internal council activities and relations with stakeholders in the school community; whether site council members feel that the council is representative of the school community; and the perceived openness of the site-based decision-making process. The four-item scale constructed for this factor is labeled *structure*; Cronbach's alpha for this scale is .81.

The final factor includes items that relate to communication and decision processes directly, items k - o on Table 1. These items relate to how council members interact with one another. Issues relating broadly to trust, fairness, and personal motives group into this factor, which is identified as *process*. Cronbach's alpha for this scale is .91.

Support

Support deals with the resources needed by site-based management teams to promote their successful operation, including time, training, stakeholder skills in collaborative decision-making, authority to make and implement decisions, and support from various stakeholder groups (see, for instance, Murphy & Beck, 1995; Wohlstetter, Smyer & Mohrman, 1994). To tap issues relating to the resources provided to stakeholders on site-based management teams, Metro respondents were asked to rate the adequacy of each of the following conditions that may affect the work of their site council (1 = very inadequate, 2 = somewhat inadequate, 3 = somewhat adequate, 4 = very adequate):

a. The time available for the site council to meet.
b. The time available to communicate with others about council decisions.
c. The time available to implement council decisions.
d. The team-building and consensus-building skills of those who facilitate or lead council meetings.
e. Council members' skills in communication and decision-making.
f. Support from building administrators.
g. Support from staff not on the council.
h. The recognition and respect site council members receive for their efforts.
i. Support from the school board.
j. Support from the superintendent.

k. Support from central administration in general.
l. The council's access to information it needs to make decisions.
m. The site council's authority to decide what issues it will address.
n. The site council's authority to implement decisions.

Relying on the descriptive literature and past work on the conditions and resources of teaching (Bacharach, Bauer, & Shedd, 1986), five factors were expected to emerge from the factor analysis: time resources (items a - c), council member skills (items d & e), support from building-level personnel (items f - h), support from higher authorities (items i - k), and the site council's decision-making authority (items l - n).

The factor analysis, which is presented in Table 2, shows that only three factors emerged. The first factor includes items relating to time, council member skills, and support from building stakeholders such as the building administrator and staff not serving on the site council (items a - h). This factor includes all items related to resources that reside at the school site; building level stakeholders and the site council can be expected to have significant influence over these resources. They are largely in the site council's control or in the control of stakeholders at the school, in contrast to resources that may be provided from the school system or an outside agent. This scale will be referred to as *council capacity*. Cronbach's alpha for this scale is .87.

Table 2
Factor Analysis of Support Items

		1	2	3
a.	The time available for the site council to meet.	**.837**	-.001	-.005
b.	The time available to communicate with others about council decisions.	**.782**	.130	.111
c.	The time available to implement council decisions.	**.719**	.132	.177
d.	The team-building and consensus-building skills of those who facilitate or lead council meetings.	**.715**	-.006	.268
e.	Council members' skills in communication and decision-making.	**.644**	-.004	.444
f.	Support from building administrators.	**.473**	.279	.387
g.	Support from staff not on the council.	**.528**	.006	.357
h.	The recognition and respect site council members receive for their efforts.	**.532**	.186	.340
i.	Support from the school board.	.164	**.911**	.168
j.	Support from the superintendent.	.121	**.916**	.144
k.	Support from central administration in general.	.005	**.864**	.197
l.	The council's access to information it needs to make decisions.	.197	.006	**.755**
m.	The site council's authority to decide what issues it will address.	.329	.263	**.715**
n.	The site council's authority to implement decisions.	.107	.370	**.709**

The second factor represents support from higher authorities, which includes items i - k on Table 2. The three-item scale will be referred to as *administrative support*; Cronbach's alpha for this scale is .92. Note that issues relating to building level support and central office support load into separate factors, and that support from building administrators loads with issues generally related to building resource issues and staff support rather than with the items dealing with support from other administrators. This may relate to the level of distrust and suspicion exhibited in Metro concerning the central office's motives in implementing site-based management. Items relating to the resources controlled at the school site (*us*) were thought of differently than those provided by the central office (*them*).

The final dimension, which includes items l - n, deals with the authority resources provided to the site-based council. Note that the survey questions dealing with access to information loads with the two items relating directly to authority. Respondents apparently recognize that real decision-making authority requires access to timely and accurate data, and that teams require access to relevant information in order to make good decisions. This three-item scale will be referred to as *authority*, and Cronbach's alpha for the scale is .76.

Outcome measures

Although Weiss (1993) observed that site-based management has sometimes been treated as an all-purpose remedy for an array of unrelated problems, there are a number of outcomes generally associated with site-based management (Conley, Schmidle & Shedd, 1988; Shedd & Bacharach, 1991). Two sets of questions were included in the Metro survey to measure outcomes typically associated with site-based management. First, respondents were asked to rate the effectiveness of site-based management in terms of the following outcomes (1=very ineffective, 2=somewhat ineffective, 3=somewhat effective, 4=very effective):

a. Providing those who are directly involved on councils influence over decision-making.
b. Providing teachers and other school staff influence over decision-making.
c. Providing parents influence over decision-making.
d. Promoting cooperation and trust among administrators and school staff.
e. Resolving problems that affect teaching, working, and learning conditions.
f. Enhancing the quality of decisions made in the school.
g. Promoting innovation.
h. Satisfying public expectations for reform.
i. Improving the education that students receive.

These items include many of the traditional explanations of the benefits of site-based management, such as the notion that it promotes trust and enhanced stakeholder influence, as well as explanations relating to improved quality of decisions, innovation and educational services (see, for example, Glickman, 1993; Hill, Bonan, & Warner, 1992; Malen et al., 1990b). Based on this literature, three factors were expected to emerge: the first related to enhancing stakeholder influence (items a - c); the second dealing with improving decision-making (items d - f); and the third relating to educational outcomes (items g - i). Factor analysis was run to confirm the presence of these three factors, and the results are presented in Table 3. Based on this analysis, three scales were developed: *effectiveness in enhancing influence* (Cronbach's alpha = .89), *improving decision-making* (Cronbach's alpha = .91), and *promoting educational outcomes* (Cronbach's alpha = .89).

Two additional, single-item measures were included in the survey to tap outcomes associated with site-based management. These questions deal with stakeholder satisfaction with two aspects of site-based management, and may be considered summary measures. The specific questions were: "Overall, how satisfied are you with your site-based council's performance?" and "Overall, how satisfied are you with the district's site-based management program?" These variables were rated on a ten point scale, with ten representing "very satisfied" and one representing "not at all satisfied." The first item will be referred to as *satisfaction with council*, and the second will be called *satisfaction with program*.

Table 3
Factor Analysis of Effectiveness Items

		1	2	3
a.	Providing those who are directly involved on councils influence over decision-making.	**.765**	.387	.322
b.	Providing teachers and other school staff influence over decision-making.	**.705**	.460	.341
c.	Providing parents influence over decision-making.	**.814**	.224	.292
d.	Promoting cooperation and trust among administrators and school staff.	.369	**.829**	.188
e.	Resolving problems that affect teaching, working, and learning conditions.	.272	**.715**	.515
f.	Enhancing the quality of decisions made in the school.	.313	**.744**	.494
g.	Promoting innovation.	.328	.305	**.815**
h.	Satisfying public expectations for reform.	.576	.247	**.621**
i.	Improving the education that students receive.	.363	.403	**.743**

A note on using perceptual outcome measures

Some may argue that to determine the efficacy of school-based management as a reform, it is essential to study the phenomena using more direct measures of student performance, such as aggregate average scores on standardized tests. However, there are several compelling reasons to use perceptual data in this study.

First, on theoretical grounds, this paper is focused on testing connections between factors that are purported to occur as intermediate events between stakeholder involvement and its impact on student learning. As mentioned earlier, the theory of school-based management is predicated on certain social-psychological consequences occurring along the way, and unless these occur, it is unlikely involvement will have its eventual impact on school performance. While perceptual changes cannot guarantee an impact on school performance, they are seen as a necessary precursor. Thus, the use of perceptual outcomes in this study may be seen as a choice that is consistent with the purpose of testing elements of the internal logic of action relating to school-based management rather than its ultimate consequence.

Second, the use of perceptual measures reflects an epistemological choice. As Bacharach, Bauer and Conley (1986) pointed out, this decision lies at the heart of an ongoing debate in organizational theory between those who view the organization as a reified structure and those who see it as a product of individually constructed realities. Bacharach, Bamberger, Conley and Bauer (1990) noted that phenomena such as involvement in decision making are best viewed from an evaluative frame; what is important is not the objective level of resources, practices, or outcomes, but rather individuals' assessment of the degree to which these sufficiently meet expectations. The constructivist approach fits the research questions posed in this study.

Finally, perceptual measures were selected based on a decision regarding the appropriate unit of analysis of the study. The theory connecting stakeholder involvement with school outcomes deals predominantly with the individual participant as the unit of analysis; an individual's participation yields greater involvement, trust, open communications, self-efficacy, and so on. In the end, the aggregate impact of these intermediate effects is assumed to result in increased school performance. Using school-level measures of performance would require aggregating individual perceptions about intermediate factors and thus reifying them. In any case, the focus of this study is on the empirical relationships between intermediate factors rather than on judging the overall impact of school-based management on aggregate school outcomes. Thus, an individual unit of analysis and perceptual outcome measures were chosen.

Control variables

As discussed earlier, two single-item variables were constructed to control for the effect of the relative newness of the site-based council and the type of school. First, a dummy variable was constructed which measures the "newness" of the site council; the eight site-based councils which were created three years prior to the study were coded "0" on this variable and the four schools which had councils for only one year were coded "1." This variable will be referred to as *new council*. Second, a dummy variable was included in the analysis which was coded "0" for elementary school councils and "1" for secondary school councils to control for the influence of size and administrative complexity on outcomes associated with site-based management. This measure is referred to as *secondary site*.

Results

Table 4 displays the descriptive statistics for the practice, support, and outcome variables. Table 5 displays the Pearson Product Moment Correlations for all variables in the regression analysis. Inspection of the correlation matrix reveals that many of the independent variables are highly correlated, thus raising the possibility of multicollinearity. Collinearity diagnostics were examined, and as a summary measure the Variance Inflation Factors (VIF) are reported along with the correlation coefficients on Table 5. Stevens (1996) states that the VIF indicates the degree of linear association between a factor and all remaining factors in the regression equation, and he suggests that in general, a VIF of greater than 10 should raise a concern. No VIF for the present study exceeds a value of approximately 3.5.

Results of the blocked regression analyses are displayed in Tables 6 through 10. Table 6 shows that there is a statistically significant relationship (R^2 = .66, F(3, 129) = 81.86, p<.01) among the four variables entered in the first step of the procedure, and the estimates of the regression coefficients indicate that scope and structure emerge as statistically significant predictors in the expected direction. The impact of scope on effectiveness at enhancing influence is quite large (eta^2 = .33), whereas the effect of structure is small (eta^2 = .04). Examination of the fully elaborated model shows that there is a statistically significant relationship among the independent variables and effectiveness at enhancing influence [R^2 = .69, F(8, 124) = 34.14, p<.01], and the scales relating to scope and authority emerge as statistically significant predictors, with scope having a large effect (eta^2 = .15). Adding the resource and control variables only slightly improves the model (ΔR^2 = .03), although this is a statistically significant improvement (ΔF = 2.55, p < .05). These results suggest that achieving some degree of consensus on goals and the authority site councils enjoy, along with being provided authority by district administration, contribute to the degree to which stakeholders in site-based management feel they can influence decision-making.

Table 4
Descriptive Statistics

	N	Minimum	Maximum	Mean	SD
Scope	133	1.00	4.00	2.72	.83
Structure	133	1.00	4.00	3.24	.70
Process	133	1.20	4.00	3.19	.79
Council capacity	133	1.13	4.00	2.94	.58
Administrative support	133	1.00	4.00	2.52	.86
Authority	133	1.00	4.00	2.81	.74
New council	133	.00	1.00	.35	.48
Secondary school	133	.00	1.00	.16	.37
Effectiveness at enhancing influence	133	1.00	4.00	2.76	.83
Effectiveness in improving decision-making	133	1.00	4.00	2.79	.83
Effectiveness at promoting educational outcomes	133	1.00	4.00	2.65	.78
Satisfaction with council	133	1.00	10.0	6.52	2.70
Satisfaction with program	133	1.00	10.0	5.34	2.55

Table 5
Zero Order Correlations for Variables in Regression Analysis

		v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	VIF
v1	Scope	–													3.34
v2	Structure	.62	--												2.47
v3	Process	.70	.70	–											3.22
v4	Council Capacity	.70	.65	.75	--										2.95
v5	Admin. Support	.38	.20	.24	.23	--									1.35
v6	Authority	.72	.55	.49	.59	.47	–								2.62
v7	New Council	**-.17**	-.25	-.19	-.17	**.00**	**-.07**	–							1.25
v8	Sec. School	**-.01**	**-.14**	-..	**-.10**	**-.14**	**-.03**	-.32	–						1.25
v9	Effective- Influence	.79	.62	.66	.66	.35	.69	-.22	**-.05**	–					
v10	Effective-Decision-making	.76	.65	.73	.69	.34	.61	-.18	**.03**	.77	–				
v11	Effective-Education Outcomes	.81	.58	.62	.66	.26	.61	-.22	**.00**	.80	.80	–			
v12	Satisf. w/ Council	.79	.62	.75	.71	.26	.50	-.25	**.02**	.68	.78	.75	–		
v13	Satisf. w/Program	.68	.36	.50	.46	.54	.55	**-.01**	**-.04**	.58	.59	.60	.59	–	

NOTE: All zero-order correlations are statistically significant (p<.05), with the exception of those in bold (p>.05).

Table 6
Regression Results - Effectiveness at Enhancing Influence (N=133)

	Step 1			Step 2		
Variable	B (s.e.)	β	η^2	B (s.e.)	β	η^2
Scope	.60 (.08)	.60••	.33	.42 (.09)	.42••	.15
Structure	.20 (.09)	.17•	.04	.09 (.09)	.08	.01
Process	.12 (.09)	.12	.02	.14 (.10)	.14	.02
Council Capacity				.08 (.12)	.05	.00
Admin. Support				.02 (.06)	.02	.00
Authority				.26 (.09)	.23••	.06
New Council				-.16 (.10)	-.09	.02
Secondary site				-.09 (.13)	-.04	.00
	R^2: .66 F: 81.86••			R^2: .69 F: 34.14•• ΔR^2: .03 ΔF: 2.55•		

• $p<.05$
•• $p<.01$

Table 7
Regression Results - Effectiveness at Improving Decision-Making (N=133)

	Step 1			Step 2		
Variable	B (s.e.)	β	η^2	B (s.e.)	β	η^2
Scope	.46 (.07)	.45••	.23	.35 (.09)	.35••	.10
Structure	.18 (.09)	.15•	.03	.18 (.10)	.15•	.03
Process	.33 (.09)	.31••	.10	.28 (.10)	.27••	.07
Council Capacity				.15 (.12)	.11	.01
Admin. Support				.07 (.06)	.08	.01
Authority				.06 (.09)	.05	.00
New Council				.04 (.10)	.02	.00
Secondary site				.20 (.13)	.09	.02
	R^2: .67 F: 87.39••			R^2: .69 F: 33.96•• ΔR^2: .02 ΔF: 1.30		

• $p<.05$
•• $p<.01$

Table 8
Regression Results - Promoting Educational Outcomes (N=133)

	Step 1			Step 2		
Variable	B (s.e.)	β	η^2	B (s.e.)	β	η^2
Scope	.67 (.07)	.71••	.41	.62 (.09)	.66••	.29
Structure	.13 (.08)	.12	.02	.08 (.09)	.07	.01
Process	.04 (.08)	.04	.00	-.02 (.09)	-.02	.00
Council Capacity				.19 (.12)	.14	.02
Admin. Support				-.05 (.05)	-.05	.01
Authority				.05 (.09)	.04	.00
New Council				-.10 (.09)	-.06	.01
Secondary site				.02 (.12)	.01	.00
	R^2: .66 F: 84.79••			R^2: .68 F: 32.61 ΔR^2: .01 ΔF: 1.10		

• $p<.05$
•• $p<.01$

Table 9
Regression Results - Satisfaction with Council (N=133)

	Step 1			Step 2		
Variable	B (s.e.)	β	η^2	B (s.e.)	β	η^2
Scope	1.64 (.23)	.50••	.29	1.86 (.28)	.57••	.26
Structure	.28 (.26)	.07	.01	.32 (.29)	.08	.01
Process	1.18 (.26)	.35••	.14	.82 (.29)	.24••	.06
Council Capacity				.79 (.37)	.17•	.03
Admin. Support				.10 (.17)	.03	.00
Authority				-.71 (.28)	-.19•	.05
New Council				-.34 (.30)	-.06	.01
Secondary site				.27 (.39)	.04	.00
	R^2: .70 F: 100.52••			R^2: .73 F: 41.38 ΔR^2: .03 ΔF: 2.47•		

• $p<.05$
•• $p<.01$

Table 10
Regression Results - Satisfaction with Program (N=133)

	Step 1			Step 2		
Variable	B (s.e.)	β	η^2	B (s.e.)	β	η^2
Scope	2.14 (.28)	.70••	.31	1.71 (.33)	.55••	.18
Structure	-.57 (.34)	-.16	.02	-.45 (.34)	-.12	.01
Process	.39 (.33)	.12	.01	.49 (.34)	.15	.02
Council Capacity				-.16 (.44)	-.04	.00
Admin. Support				.94 (.20)	.32••	.15
Authority				.09 (.33)	.03	.00
New Council				.44 (.35)	.08	.01
Secondary site				.12 (.45)	.02	.00
	R^2: .48 F: 39.70••			R^2: .58 F: 21.23•• ΔR^2: .10 ΔF: 5.74••		

• $p<.05$
•• $p<.01$

Table 7 displays the results for the regression equations dealing with effectiveness at improving decision-making. Inspection of step one shows that there is a statistically significant relationship [$R^2 = .67$, $F(3, 129) = 87.39$, p<.01] among the variables entered, and the estimates of the regression coefficients indicate that scope, structure, and process emerge as statistically significant predictors in the expected direction. Scope has a large effect on the dependent measure ($\text{eta}^2 = .23$), process a moderate effect ($\text{eta}^2 = .10$), and structure has a small effect ($\text{eta}^2 = .03$). Examination of step two of the procedure shows that there is a statistically significant relationship among the independent variables and effectiveness at improving decision-making [$R^2 = .69$, $F(8, 124) = 33.96$, $p<.01$)], and each of the factors measuring decision-making and communication practices emerge as statistically significant in the expected direction. Adding the resource and control variables does not improve the model ($\Delta R^2 = .02$, $\Delta F = 1.30$, $p > .05$), and none of the resource or control variables are statistically significant. These results imply that the decision-making and communication practices engaged in by site council members are important to the degree to which participants feel that site-based management can improve decision-making practice in their school.

The results for the regression analysis dealing with effectiveness at improving educational outcomes is shown in Table 8. Step one shows that there is a statistically significant relationship [$R^2 = .66$, $F(3, 129) = 84.79$, $p < .01$] among the variables entered, and the estimates of the regression coefficients indicate that scope is a statistically significant predictor and has a large effect on the dependent variable ($\text{eta}^2 = .41$). Examination of step two of the procedure shows that there is a statistically significant relationship among the independent variables and effectiveness at improving decision-making [$R^2 = .68$, $F(8, 124) = 32.61$, $p < .01$)], although adding the resource and control variables does not improve the model ($\Delta R^2 = .01$, $\Delta F = 1.10$, $p > .05$). The only statistically significant predictor in this equation is scope, and it has a large effect on improving educational services ($\text{eta}^2 = .29$), implying that issues like goal consensus and providing site councils with a degree of influence on issues of importance are key factors in ensuring that site-based management has an impact on teaching and learning.

Tables 9 and 10 display the results for the regression equations dealing with the two satisfaction measures. Inspection of step one on Table 9 shows that there is a statistically significant relationship [$R^2 = .70$, $F(3, 129) = 100.52$, $p < .01$] among the practice variables and satisfaction with the site council. Estimates of the regression coefficients indicate that scope and process emerge as statistically significant predictors in the expected direction, with both variables having a large effect on the dependent measure ($\text{eta}^2 = .29$ for scope, .14 for process). Step 2 shows that there is a statistically significant relationship among the independent variables and satisfaction with the site council [$R^2 = .73$, $F(8, 124) = 41.38$, $p < .01$)], and adding the resource and control variables makes a small but statistically significant contribution ($\Delta R^2 = .03$, $\Delta F = 2.47$, $p < .05$). Scope and process remain as statistically significant predictors, with scope having a large effect size ($\text{eta}^2 = .26$) and process having a moderate effect ($\text{eta}^2 = .06$). Council capacity also emerges as a statistically significant predictor in the expected direction, although its effect is small ($\text{eta}^2 = .03$). Authority emerges as statistically significant with a moderate effect on the dependent variable ($\text{eta}^2 = .05$), although the relationship is in the opposite direction than hypothesized, implying that the more adequate respondents felt their authority was, the lower their satisfaction with the site council. Given the fact that councils in Metro felt stalled and were generally suspicious concerning the central office's motives, it may be that authority taps into expectations for performance, and thus the more authority stakeholders felt their council had, the less pleased they felt with the progress the council was making.

Table 10 shows that there is a statistically significant relationship among the practice variables and satisfaction with the district program, although the magnitude of this relationship is considerably smaller than in the previous equations [$R^2 = .48$, $F(3, 129) = 39.70$, $p < .01$]. Estimates of the regression coefficients indicate that scope emerges as a statistically significant predictor in the expected direction, and it has a large effect ($\text{eta}^2 = .31$) on the dependent variable. Step 2 shows that there is a statistically significant relationship among the independent variables and satisfaction with the site council [$R^2 = .58$, $F(8, 124) = 21.23$, $p < .01$)], and adding the resource and control variables in this case makes a large and statistically significant contribution ($\Delta R^2 = .10$, $\Delta F = 5.74$, $p < .01$). Scope remains as a statistically significant predictor with a large effect on the dependent variable ($\text{eta}^2 = .18$), and administrative support emerges as a statistically significant predictor with a large effect ($\text{eta}^2 = .15$). It is not surprising that this variable has a large effect; this implies that the more stakeholders feel that the board, superintendent, and central office provide adequate support to site councils, the more satisfied participants are with the district program.

Summary

The purpose of this paper was to investigate the relationship between factors relating to the decision-making and communication practices engaged in by participants in site-based management and perceived outcomes associated with this reform. Two research questions were posed, the first asking whether there is a relationship between measures relating to communication and decision-making practices and perceived outcomes, and the second asking whether any empirical relationship between these factors would hold up when accounting for the effect of measures relating to factors most often cited by researchers as causally related to the efficacy of site-based management, the resources provided to site team. Four hypotheses were tested using blocked regression analysis to investigate these research questions.

Table 11 presents a summary of the results of the regression analysis. For each dependent measure, statistically significant predictors are shown along with the effect size for each of these factors. Overall, there is considerable support for H_1, which predicted that scope, structure, and process would have a statistically significant, positive impact on perceived outcomes. Note that in all but the final equation, a large percentage of the variance in the dependent measures is accounted for by the practice factors alone (R^2 ranges from .66 to .70 in step one). Scope emerges as statistically significant in step two of each model, and has a large effect on the dependent measure in all but one case. This implies that regardless of outcome associated with site-based management, it is important to achieve a consensus on goals and the amount of authority delegated to site teams, and participants must feel that they have influence over important decisions. Process is important with respect to the effectiveness of site-based management at improving decision-making and the degree of satisfaction stakeholders experience with their site council, suggesting that issues related to communication and trust are important elements to promoting these outcomes.

In terms of the aggregate contribution of the resource and control variables, there is little support for H_2, which stated that we expected the resource scales to have a statistically significant, positive relation to perceived outcomes. The blocked regressions show that adding the resource and control variables has only a very slight effect, if any, except with regard to the degree of satisfaction stakeholders experience with the district program. Administrative support emerges as a statistically significant predictor for this outcome, and has a large effect, but the other resource scales emerge only a few times and have only small or medium effects on the outcome measures. Likewise, there is no support for H_3 and H_4, indicating that the two control variables have little impact on the perceived outcomes of site-based management.

Discussion

The findings from this study have implications for both researchers and practitioners. For researchers interested in studying school reform and site-based management, findings suggest strongly that it is necessary to account for decision-making and communication practices when conducting research into the efficacy of site-based management. To date, the literature has not adequately dealt with the actual processes site-based teams use in their decision-making and planning practice, and there are few empirical analyses of elements of site-based practice that concern how teams actually operate. Site-based practice is deemed "too complex" to operationalize (Malen et al., 1990b; Wohlstetter & Odden, 1992), and thus there are few explications of the specific variables that are important to the success of site-based management. This study suggests, however, that to fully understand the promise of site-based management and the complex connections that occur between involvement and various outcomes, it is critical to develop models that deal specifically with variables related to the practice of site-based management.

The measures dealing with scope, structure, and process presented here represent a start. A useful next step would be for researchers to use these measures in studies in other districts, both to expand our understanding of the relationship of factors relating to practice and to determine if the findings presented here are generalizeable to other sites. Metro is a single, urban district; the results and conclusions should be applied to other settings with caution. Additional measures need to be constructed, as well; the factors presented here are fairly general, and they are hardly inclusive of all elements of practice involved in site-based management. If an understanding of the complexities of site team practice is to be developed, much more detailed articulation of the these phenomena are needed.

Table 11
Summary of Results

	Enhancing Influence	Improving Decision-Making	Improving Educational Outcomes	Satisfaction with Council	Satisfaction with Program
Scope	** / .15	** / .10	** / .29	** / .26	** / .18
Structure		* / .03			
Process		** / .07		** / .06	
Council				* / .03	
Capacity					** / .15
Admin. Support	** / .06			-* / .05	
Authority					
New Council					
Secondary site					

Note: Statistically significant regression coefficients from step 2 of the regression analyses are displayed (*$p < .05$, **$p < .01$), along with the effect size (partial eta^2).

It will also be necessary to develop more fully elaborated theoretical models that deal with such issues as the relationships among elements of site-based practice and factors like resources. The absence of theoretical models that deal with the relationships among inputs, practice, and outcomes limits the scope of investigation. For instance, the findings presented in this paper suggest that the adequacy of resources provided to site teams is relatively unimportant to various outcomes. However, it seems unlikely that resources would be largely unrelated to perceived outcomes. It is possible that either the practice and resource factors are related to common unmeasured antecedent variables, or that they are related hierarchically (e.g., resources influence the communication and decision-making practices in some way, and these in turn influence outcomes). In any case, there is little theory to help us understand the nature of the relationships among input factors like resources, site team practices, and outcomes. The preliminary nature of the present study, the absence of theory dealing specifically with elements of practice, and the lack of a research base to establish hypotheses about an appropriate hierarchical or path model, reflect a gap in the literature that scholars need to address.

The findings have implications for practitioners interested in successfully implementing site-based management, as well. Specifically, findings suggest that the district must play a more significant role in site-based change management. Like many districts nationwide (McGonagill, 1993), the parameters within which site-based teams were expected to operate were left ambiguous in Metro. Questions like the degree of authority teams had and the types of issues they could address were considered an infringement on school-site council autonomy. Bacharach (1990) observed that an organization that is unclear about its goals is incapable of strategic reform; similarly, Taylor and Levine (1991) suggested that without a clearly articulated reason for pursuing site-based management, changes in planning and decision-making arrangements will result in only cosmetic changes. Implementing a process like site-based management without clearly articulating its purpose and the district's expectations regarding site team performance risks the process being seen as an end in itself (Murphy & Beck, 1995; NASSP, 1992).

There has been relatively little mention in the restructuring literature on the role of the district in supporting school-based changes save for consistent calls for adequate resources (Bauer, Meza & Duplaintis, 1999). However, given the importance of the decision-making and practice factors in this study, one conclusion that may be drawn is that an equally important responsibility of the district is to provide a clear picture of the goals and processes of site-based management. For example, the factor measuring scope involves such issues as whether site council members and district authorities agree on decision-making power, the clarity of goals, and whether the site council has real influence on important decisions. These issues relate to the design of site-based management and the policy governing the practice in a given district. District officials must clarify the parameters within which site teams work, including such things as the teams' decision-making influence and the goals for which teams will be held accountable. Similarly, findings suggest that school leaders need to work to create a context in which collaborative communication and decision-making processes can thrive, issues that are embedded in the process factor.

This study stems from the simple notion that the nature and quality of decision-making and communication practices experienced by stakeholders in site-based management is important to valued outcomes. The analysis suggests an important conclusion: If site-based management is to be more than an end in itself, and if the process is going to contribute in a meaningful way to school improvement, site-based teams need more than adequate resources. They need the leadership and support to develop common, realistic expectations and a sense of the limits of site-based practice in their district. Furthermore, if research is going to contribute meaningfully to practice, more research needs to be conducted on the factors relating to the steps that occur between stake-holder involvement and school improvement, including issues relating to communication and decision-making practice.

References

Adams, J. (1988). The role of the creative outlook in team building. In W. Reddy (Ed.), *Team building: Blueprints for productivity and satisfaction* (pp. 98-106). Alexandria, VA: NTL Institute for Applied Behavioral Science and San Diego, CA: Pfeiffer and Company.

Akin, J. (1992). *When the restructuring bandwagon reaches high school, will site culture bar the door?* [On-line]. Available: ERIC 352728.

Bacharach, S. (1990). Education reform: Making sense of it all. In S. Bacharach (Ed.), *Education Reform: Making Sense of It All*. Boston: Allyn and Bacon.

Bacharach, S., Bamberger, P., Conley, S., & Bauer, S. (1990). The dimensionality of decision participation in educational organizations: The value of a multi-domain evaluation approach. *Educational Administration Quarterly, 26*(2), 126-167.

Bacharach, S., Bauer, S., & Conley, S. (1986). Organizational analysis of stress: The case of elementary and secondary schools. *Work and Occupations, 13*(1), 7-32.

Bacharach, S., Bauer, S., & Shedd, J. (1986). The work environment an school reform. *Teachers College Record, 88*(2), 241-256.

Bauer, S. (1998). Designing site-based systems, deriving a theory of practice. *International Journal of Educational Reform*, 7(2), 108-121.

Bauer, S., Meza, J., & Duplaintis, M. (1999). *Building a Conceptual model for shared accountability in an accelerated school district*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Bettenhausen, K. (1991). Five years of groups research: What have we learned and what needs to be addressed. *Journal of Management, 17*(2), 345-381.

Burke E.W. (1988). Team building. In W. Reddy (Ed.), *Team Building: Blueprints for Productivity and Satisfaction* (pp. 3-14). Alexandria, VA: NTL Institute for Applied Behavioral Science and San Diego: Pfeiffer & Company.

Cohen, M. (1988). *Restructuring the educational system: Agenda for the 1990's.* Washington, DC: National Governor's Association.

Conley, S. & Bacharach, S. (1987). The Holmes group report: Standards, hierarchies, and management. *Teachers College Record, 88*(3), 340-347.

Conley, S., Schmidle, T., & Shedd, J. (1988). Teacher Participation in the management of school systems. *Teachers College Record, 90*(2), 259-280.

Corcoran, T., Hansen, B., & Shidlowski, A. (1988). *Who makes decisions? The role of New Jersey teachers in school decision-making.* Trenton, NJ: The New Jersey School Boards Association.

Cotton, J. (1993). *Employee involvement: Methods for improving performance and work attitudes.* Newbury Park, CA: Sage Publications.

David, J. (1989). Synthesis of research on school-based management. *Educational Leadership, 46*(9), 45-53.

Douglass, M. & Douglass, D. (1992). *Time management for teams.* New York: AMACOM.

Glickman, C. (1990). Pushing school reform to a new edge: The seven ironies of school empowerment. *Phi Delta Kappan, 72*(1), 68-75.

Glickman, C. (1993). *Renewing America's schools: A guide for school-based action.* San Francisco: Jossey-Bass Publishers.

Guzzo, R. & Dickson, M. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology, 47*, 307-338.

Hatry, H., Morley, E., Ashford, B., & Wyatt, T. (1993). *Implementing school-based management: Insights into decentralization from science and mathematics departments.* Washington, DC: The Urban Institute.

Hill, P., Bonan, J., & Warner, K. (1992). Uplifting education: Set schools soaring with site-based management. *American School Board Journal, 179*(3), 21-25.

Johanson, G., Green, S., & Williams, J. (1999). Blunders and missed opportunities in the analysis and interpretation of survey data. *The Respondent* (Summer), 7-9.

Kanter, R. (1983). *The change masters: Innovation for productivity in the American corporation.* New York: Simon and Schuster.

Katzenbach, J. & Smith, D. (1993). *The wisdom of teams: Creating the high performance organization.* Boston: Harvard Business School Press.

Kirby, P. (1992). Shared decision making: Moving from concerns about restrooms to concerns about classrooms. *Journal of School Leadership, 2*(3), 330-344.

Lindquist, K. & Mauriel. J. (1989). School-based management: Doomed to failure. *Education and Urban Society, 21*(4), 403-416.

Malen, B. & Ogawa, R. (1992). Site-based management: Disconcerting policy issues, critical policy choices. In J. Lane & E. Epps (Eds.), *Restructuring the schools: Problems and prospects* (pp. 185-206). Berkeley, CA: McCutchan Publishing Corp.

Malen, B., Ogawa, R., & Kranz, J. (1990a). Unfulfilled promises: Evidence says site-based management hindered by many factors. *The School Administrator, 47*(2), 30-32; 53-56, 59.

Malen, B., Ogawa, R., & Kranz, J. (1990b). What do we know about school-based management? A case study of the literature - a call for research. In W. Clune & J. Witte (Eds.), *Choice and control in American education, volume 2: The practice of choice, decentralization and school restructuring* (pp. 289-342). London: The Falmer Press.

Marsh, D. (1994). Change in schools: lessons from the literature. In S. Mohrman, P. Wohlstetter, & associates (Eds.), *Designing high performance schools: Strategies for school-based management* (pp. 215-252). San Francisco: Jossey-Bass Publishers.

McGonagill, G. (1993). *Overcoming barriers to educational restructuring: A call for system literacy.* Arlington, VA: American Association of School Administrators.

Miles, M. & Louis, K. (1990). Mustering the will and skill for change. *Educational Leadership, 47*(8), 57-61.

Murphy, J. & Beck, L. (1995). *School-based management as school reform: Taking stock.* Thousand Oaks, CA: Corwin Press.

National Association of Secondary School Principals. (1992). *A Leader's Guide to School Restructuring: A Special Report of the NASSP Commission on Restructuring*. Reston, VA: Author.

Ogawa, R. & White, P. (1994). School-based management: an overview. In S. Mohrman, P. Wohlstetter, & associates (Eds.), *Designing high performance schools: Strategies for school-based management* (pp. 53-80). San Francisco: Jossey-Bass Publishers.

Parker, G. (1990). *Team players and teamwork: The new competitive business strategy*. San Francisco: Jossey-Bass Publishers.

Parker, G. (1994). *Cross-functional teams: Working with allies, enemies and other strangers*. San Francisco: Jossey-Bass Publishers.

Robbins, H. & Finley, M. (1995). *Why teams don't work*. Princeton, NJ: Peterson's / Pacesetter Books.

Sharpe, F. (1996). Towards a research paradigm on devolution. *Journal of Educational Administration*, *34*(1), 4-23.

Shedd, J. (1987). *Involving teachers in school and district decision-making*. Ithaca, NY: Organizational Analysis and Practice, Inc.

Shedd, J. & Bacharach, S. (1991). *Tangled hierarchies: Teachers as professionals and the management of schools*. San Francisco: Jossey-Bass Publishers.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Taylor, D. & Bogotch, I. (1994). School-level effects of teachers' participation in decision-making. *Educational Evaluation and Policy Analysis*, *16*(3), 302-319.

Taylor, B. & Levine, D. (1991). Effective schools projects and school-based management. *Phi Delta Kappan*, *72*(5), 394-397.

Weiss, C. (1993). Shared decision making about what? A comparison of schools with and without teacher participation. *Teachers College Record*, *95*(1), 69-92.

Wohlstetter, P. & Odden, A. (1992). Rethinking school-based management policy and research. *Educational Administration Quarterl,*, *28*(4), 529-549.

Wohlstetter, P., Smyer, R. & Mohrman, S. 1994. New boundaries for school-based management: The high involvement model. *Educational Evaluation and Policy Analysis*, *16*(3): 268-286.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 1, 29-43

An Initial Reintegration Treatment of Children with Acute Lymphoblastic Leukemia (ALL)

Michelle Lurie
Fulbright & Associates, Dallas, TX

Nadeen Kaufman
Yale University School of Medicine

The purpose of this study was to evaluate the cognitive, psychological and social adjustment of pediatric acute lymphoblastic leukemia (ALL) patients and to assess how their needs could best be met by means of appropriate reintegration programs. These reintegration programs focused on the learning/educational needs of the children, as well as their psychosocial functioning. Three qualitative case studies of pediatric ALL patients were conducted. Results of psychoeducational assessments administered prior to reintegration indicated nonspecific attention problems, particularly with regards to auditory attention. Planning and organizational difficulties, along with some memory problems were also evident. These deficits, along with social and/or emotional difficulties, were addressed in the reintegration programs. These reintegration programs were tailored to meet the unique cognitive and psychosocial needs of each individual participant, with the aim of providing each child with appropriate services and accommodations. This study highlights the need for each ALL patient to be provided with a comprehensive program integrating medical, social, psychological, and educational components.

Introduction

Cancer is the leading cause of death due to illness in childhood and adolescence, except in infancy. At the same time, advances in the treatment of childhood cancer have resulted in projected survival rates of 80% for children diagnosed in 1990. The most common form of childhood cancer is acute lymphoblastic leukemia (ALL) which accounts for approximately 30% of all new cancer-diagnosed children (Mulhern, Ochs, & Fairclough, 1992). The term ALL refers to a group of heterogeneous diseases in which there is a malignancy of the bone marrow that produces blood cells. There are about 2,000 new cases of ALL diagnosed per year in the United States, or about 4 per 100,000 children younger than 15 years of age (Poplack, 1989).

The increasing prevalence of survivors of ALL has had several major impacts on the health care, family, and education systems. Children with ALL and other chronic conditions use more and a wider array of health services than other children. They also make up an increasing proportion of the practice of the primary care pediatrician, particularly as acute diseases of childhood have declined. However, to have a chronic illness is a state of life. Therefore, medical attention is not enough. Furthermore, pediatricians are rarely equipped to help the child with ALL beyond his or her physical needs. Parents are often also unprepared for this task, particularly since chronic childhood conditions increase the risk for parental adjustment difficulties, often resulting in functional limitations in the child (Thompson & Gustafson, 1995). Schools, too, face a problem never foreseen. The methodology to teach the normal child or a child with learning disabilities might not suit the needs of the child with ALL due to a variety of cognitive, emotional and psychosocial difficulties that often accompany the diagnosis of ALL. The needs of the child with ALL are therefore often well beyond that which can be provided by means of existing special education programs. Since neither the pediatrician, parent, or teacher can individually help the child with ALL with all of his or her needs, the psychologist has a unique role to play in negotiating between the medical, family, and education systems to best meet the needs of pediatric ALL patients.

Neuropsychological and Psychosocial Difficulties

Clearly, while the ultimate goal of cancer therapy is patient survival, medical survival is not always without cost to the patient. Despite the medical advances in the treatment of chronic illnesses, children with a chronic health condition have long been considered at excess risk for neuropsychological and psychosocial difficulties.

The first author conducted this research while at the California School of Professional Psychology in San Diego as part of her doctoral dissertation. Michelle Lurie is a provisionally licensed psychologist at Fulbright & Associates, Dallas, Texas; Nadeen Kaufman is a Lecturer on the Clinical Faculty of the Yale University School of Medicine. Please address correspondence regarding this manuscript to Michelle Lurie, 18735 Park Grove Lane, Dallas, TX 75287 or by e-mail to drshel@flash.net.

Research has investigated the effect of central nervous system (CNS) prophylaxis on neuropsychological development in children with cancer; however, the results of this literature tend to be confusing and inconsistent. The earliest study of intellectual deficits resulting from ALL and associated therapies reported that pediatric ALL patients who had received CNS treatments were essentially free of major neuropsychological and psychological difficulties (Soni, Marten, Pitner, Duenas, & Powazek, 1975). More recent and methodologically rigorous studies have indicated that CNS prophylaxis results in deficits in intellectual ability, memory, sustained attention, distractibility, and academic achievement (Stebhens et al. 1991). Because of the increasing interest in this area and the contradictory findings, several authors have reviewed the existing literature to attempt to resolve the differences. Williams and Davis (1986), in their summary of previous research in this area, conclude that the majority of children with ALL do not suffer severe or even moderate intellectual deficits. In contrast, Fletcher and Copeland (1988) reviewed 41 studies of the effects of prophylactic CNS treatment in children with cancer. They concluded that despite significant differences in research design, sample size, and outcome variables, the studies suggest that CNS prophylaxis does impair cognitive development, particularly when cranial radiation therapy (CRT) is part of the treatment. In another review of published studies, Cousens, Waters, Said, and Stevens (1988) reported a meta-analysis of 20 studies of children with ALL. This review revealed that CRT results in a Full Scale IQ decrement of about 10 points. Younger age at the time of CNS radiation and time elapsed since treatment both contributed to this overall effect. The type of control group used in the reviewed studies also impacted the results of the study. The average ALL patient's FSIQ was lower than 81% of healthy controls and 72% of nonirradiated cancer controls. Cousens et al. therefore concluded that the experience of pediatric cancer alone, independent of radiation treatment, results in a lowering of IQ. However, they noted that the size of this effect is smaller than the effect of direct CNS treatment.

It is therefore clear, as concluded by Stebhens et al. (1991), that writers who review and synthesize the collective findings of individual studies differ in their interpretation of the effects of CNS prophylaxis on FSIQ. However, it appears that primary dependence on IQ as the measure of neuropsychological integrity may be a problem since it is a global measure and may not be immediately sensitive to some important changes in brain function. It is therefore critical to investigate the integrity of the different components of neuropsychological functioning in children with ALL. These components may include memory and learning, attention, language, visual-spatial and perceptual-motor abilities.

Deficits in the performance of ALL patients on these neuropsychological components have been reported by Brouwers, Riccardi, Poplack, and Fedio (1984) (attention); Peckham, Meadows, Bartel, and Marrero (1988) (visual and auditory memory); Stebhens et al. (1991) (expressive language functions); Meadows et al. (1981) and Moore, Kramer, Wara, Halberg, and Ablin (1991) (visual-motor integration); Whitt, Wells, Lauria, Wilhelm, and McMillan (1984) and Ciesielski et al. (1994) (visual-motor coordination). Therefore, any assessment of the neuropsychological functioning of ALL children should include an evaluation of these components.

The emotional and behavioral functioning of pediatric cancer patients have also been researched. Here, too, conflicting findings have been reported. According to Noll, Bukowski, Rogosh, LeRoy, and Kulkarni (1990), reports by teachers suggest that children with cancer have fewer leadership and positive social skills, are disengaged from peers, and have difficulty coping with daily academic and interpersonal classroom demands. These children are perceived by their teachers as being less socially competent and as more socially isolated. This seems to indicate that children with cancer do have real problems adjusting to the social, behavioral, and academic demands of the classroom, and have difficulties with day-to-day functioning. Contrary to these findings and those of other studies, in a recent study Noll et al. (1999) report that children with cancer currently receiving chemotherapy were functioning better socially and similarly emotionally to case controls. Relative to these controls, the children with cancer were perceived as being more sociable, less aggressive, and having greater social acceptance. Measures of depression, loneliness, anxiety, and self-concept showed no significant differences, except children with cancer reported significantly lower satisfaction with current athletic competence. Noll et al. (1999) hypothesize that this lowered athletic self-concept may be due to lower energy and chronic fatigue when receiving chemotherapy. When discussing these results, the researchers note that the repeated exposure to stressful events in the context of support from parents and medical staff may strengthen the child. They also note that the experience of cancer and the possibility of death, along with painful treatments and side effects from chemotherapy, may increase sensitivity to the needs of others and lead to better social functioning. In addition, the fatigue or general malaise that may accompany treatment may lead to the child being less aggressive and therefore perceived in more positive ways within social domains. Noll et al. (1999) add that only if there is severe interference with normative parental functioning, such as a lack of monitoring or excessive coercion, or if the child experiences damage to the CNS, are problems in the social and emotional development of these children noted.

Stebhens et al. (1991) note that while some studies have indicated significant problems in self-esteem, depression, and anxiety, others have not found problems in any of these areas. These researchers comment that the conflicting findings are possibly the results of several methodological and design variables, such as the source of information, the use of non-standardized measures, the lack of blind data collection procedures, and the absence of longitudinal data. Madan-Swain and Brown (1991) report that the measures used to examine depression in pediatric cancer survivors are not particularly sensitive to the unique socialization and adjustment difficulties experienced by these children. According to these researchers, much of the research on children with ALL has suggested that these children have problems associated with lethargy, poor concentration, and low self-esteem. Although these symptoms may be indicative of depression, these researchers caution that they may also be a side effect of chemotherapy. They therefore recommend further investigation to determine the relative contributions of children's social experiences, biological changes due to the disease, and associated chemotherapeutic treatments.

The pediatric cancer patient's psychological and social adjustment may also be negatively impacted by the visible side effects of treatment, such as hair loss and weight gain or loss. In our society, physical attractiveness represents a highly prominent personal characteristic, which influences interpersonal interactions as well as perceptions about the abilities of others (Varni, 1993). Given society's attitude towards physical appearance, it may be expected that children with visible physical differences would experience discrimination in their social environment. Recently, Varni delineated "perceived physical appearance" as a predictor variable within a multivariate conceptual model to explain in part the observed variance in adaptation by children with physical differences. It might be expected that the age of the child would impact the strength of this predictor variable. Katz (1980) notes that while the preschool child may rarely experience baldness or weight gain as a severe disability due to his or her lack of preoccupation with appearance, the adolescent often experiences physical changes with the highest level of anxiety, since at this age personal appearance and peer acceptance are of primary importance. According to Ross and Ross (1984), children with leukemia report that the experience of being teased on returning to school because of hair loss and weight changes is worse than the physical pain from the disease or diagnostic and treatment procedures. Similarly, Katz and Varni (1993) note that children with newly diagnosed cancer are more likely to display a "social vulnerability" since they must endure highly visible physical appearance changes, all of which can result in significant negative reactions from peers. Varni, Katz, Colgrove, and Dolgin (1995) report that children with cancer who experience disease and treatment-related changes in physical appearance are more likely to experience depressive symptoms, social anxiety, and lower general self-esteem. It appears that children with ALL are at some risk for increased psychosocial problems. However, not all children with ALL have mental health problems. This highlights the need for research and better understanding of the coping strategies used by children with ALL in the face of their different life experiences.

School Difficulties

In addition to the neuropsychological and psychosocial impact of ALL on the child, ALL also impacts the child's attendance and performance at school. Research has examined pediatric cancer patients' school experiences and absence rates (Brown & Madan-Swain, 1993; Cairns, Klopovich, Hearne, & Lansky, 1982; Charlton et al., 1991; Katz, Varni, Rubenstein, Blew, & Hubert, 1992; Lansky, Cairns, & Zwartjes, 1983; Rynard, Chambers, Klinck, & Gray, 1998; Stebhens et al., 1991; Stehbens, Kisker, & Wilson, 1983; Taylor, Albo, Phebus, Sachs, & Bierl, 1987). These researchers have identified difficulties with school adjustment and attendance, school performance problems, and academic difficulties among childhood cancer survivors. Other researchers have examined the impact of school reentry programs on a child's adjustment (Katz, Rubenstein, Hubert, & Blew, 1988; Katz, Varni, Rubenstein, Blew, & Hubert, 1992). School reintegration programs involve cooperative efforts among health care providers, the child and family, and the school system. One of the objectives of such programs is to elicit and maintain peer support. School reentry programs are therefore cost-effective interventions because, when handled appropriately, they can prevent many future social and peer problems for the patient. They can ensure continuity of education and promote the continued acquisition of age-appropriate adaptive behavior skills (Deasy-Spinetta, 1993).

Despite the psychosocial, neuropsychological, and school difficulties experienced by children with cancer and the clear need for effective reintegration programs, Gortmaker, Walker, Weitzman and Sobol (1990) note that mental health services for these children, and children with other chronic illnesses, remain fragmented, signaling the need for increased attention to behavioral problems and their treatment among all health professionals caring for these children. Although the school system may have various programs to address the needs of chronically ill children, the efficacy of the programs is questionable if they are not tailored to individual needs. According to

Masera et al. (1995), the reintegration of the child with cancer into school is an essential part of the total treatment program. Only cooperation among the family, care team, and educators, in addition to a well-structured program, can help in achieving this goal and preventing serious negative outcomes. The psychologist can play an important role in carefully assessing the needs of the child, and drawing on the resources provided by the various systems within which the child interacts to best meet these needs.

Statement of the Problem

There are multiple studies assessing the functioning of ALL pediatric patients; however, research has not focused on the individual needs of these patients and how these needs could best be addressed in a school reentry program. The present study consisted of three qualitative case studies and aimed to assess the neuropsychological and psychosocial needs of three children with Acute Lymphoblastic Leukemia (ALL), and to explore how these needs could best be addressed by means of an appropriate reintegration program. An appropriate reintegration program is one that utilizes information from a wide variety of domains, such as educational, social, psychological, and cultural, and, on the basis of this information, provides guidance to teachers, parents, and peers as to how they can best help the patient reach his or her potential in the classroom.

Method

Subjects

Three qualitative case studies were conducted on ALL pediatric patients at the Children's Hospital of Orange County (CHOC), California. All identifying information in this article has been changed to protect the confidentiality of the patients. Two Caucasian females, Kim, aged 7½ and Susie, aged 6 years, and Peter, a 7-year-old Hispanic male, were referred to this study by the oncology/hematology team and the Health Psychology Department at CHOC. They had received CNS prophylaxis involving chemotherapy and had less than 2 years of maintenance therapy. Their treatments had not included CRT or bone marrow transplantation, and therefore none of the children had recently been exposed to anesthesia. The children referred to this study were all in private elementary schools. None of the children in this study had had any previous testing, either privately or in the schools. The consent of parents, along with the assent of the children, was obtained, and IRB approval for the study was attained.

It was preferable to select children between the ages of 6 to 8 in the early years of elementary school since this increases the possibility of each child working within a self-contained classroom and therefore having one teacher with whom he or she works most frequently. This teacher was able to provide the most information on the child's classroom functioning. Since children in the early stages of schooling work less independently than those in higher grades, the teaching style of this teacher was anticipated to affect the child's classroom performance.

Assessment Procedures

Data were collected on the patients by means of extensive neuropsychological assessment, child interview, parent interviews, behavior checklists, school observation, review of school records, a measure of the teacher's instructional style, teacher interview, review of medical records, and interview with relevant medical personnel working with the child to assess their concerns regarding the child's functioning. A major component of the neuropsychological assessment was the administration of the Developmental Neuropsychological Assessment (NEPSY; Korkman, Kirk, & Kemp, 1998). The NEPSY, an acronym from the NE in neuro- and PSY in –psychological, is a comprehensive instrument designed to assess neuropsychological development in preschool and school-age children between the ages of 3 and 12. It is based on the Lurian theoretical model and measures five functional domains: attention and executive functions, language, sensorimotor skills, visual-spatial processing, and learning and memory. Age-related means and standard deviations are available for the subtests in each domain. This allows the clinician to estimate the level at which a child's functioning within a domain is intact and the level at which a child begins to have difficulty. This information is helpful in making recommendations for intervention.

The initial assessment of each child included the administration of the NEPSY, along with a short form of the Wechsler Intelligence Scale for Children – Third Edition (WISC-III) (Wechsler, 1991). The short-form of the WISC-III used in the current study included the Similarities, Arithmetic, Picture Completion and Block Design subtests. According to Kaufman, Kaufman, Balgopal, and McLean (1996), this short form is psychometrically strong, generates clinically rich information, and is practical in that it takes about 27 minutes to administer and is quick to score. This short-form of the WISC-III provided the examiner with some range within which the child would be expected to perform on all cognitive tests. In addition to these tests, the initial testing session also included the Kinetic Family Drawing (KFD), human figure drawings, and a sentence completion task. The parents of each child were asked to complete the Barkley Developmental History Form (Barkley, 1990) to obtain a record of the child's development and previous social or academic difficulties. The parents were also asked to complete the

Child Behavior Checklist (CBCL) (Achenbach, 1991a). Each parent individually completed a checklist so as to obtain his or her own perspective of the child's functioning.

After the initial assessment of each child, at least one additional testing session was scheduled during which the neuropsychological assessments were individualized to obtain a comprehensive understanding of each child's cognitive and emotional functioning, depending upon the child's performance on earlier tests. The child was also observed at school using the Student Observation System (SOS), a component of the Behavior Assessment System for Children (BASC) (Reynolds & Kamphaus, 1992). The child's teacher was asked to complete the Teacher Report Form (TRF) (Achenbach, 1991b) to obtain a measure of his or her view of the child's performance in class. The TRF is the teacher-completed analogue to the CBCL. In order to assess the teaching style of the teacher, the Instructional Styles Inventory (ISI) (Canfield & Canfield, 1988) was administered. This measure of the teacher's teaching style was useful in designing a school reintegration program, since the researcher was able to assess whether the instructional style was well matched with the needs of the student as determined by the comprehensive psychoeducational assessment. This information was then used in recommending productive teaching techniques to help the child in his or her adjustment back into the regular classroom environment. A complete list of the tests administered, the components of the school visit, and other possible tasks performed depending upon the needs of the child are outlined in Table 1. The data collected were combined to assess the educational and psychosocial needs of each participant in the study. This information was then utilized in creating for each child an individualized reintegration program aimed at providing the child with appropriate services and accommodations. Each program combined the efforts of the school, parents, and peers to assist the children in re-adapting to school.

A feedback session was scheduled with the parents of each participant during which time they were provided with a verbal description of their child's cognitive and emotional functioning, along with a comprehensive written report of the assessment. They were also advised as to the most appropriate reintegration program which would best meet their child's needs. The researcher also met with each child's teacher to further advise them on the appropriate services and accommodations necessary for effective reintegration of the child into the school. For each participant in the study, the researcher conducted a teacher interview which aimed at assessing the teacher's knowledge about and emotional reactions to cancer. The researcher also performed a classroom presentation for the patient's classmates to provide the children with age-appropriate information about leukemia and its treatments. The effectiveness of this presentation was assessed by means of a newly constructed, brief questionnaire that was given to the class before the presentation and again at its conclusion (see Table 2). This allowed the researcher to determine whether the presentation was successful in dispelling common myths about the disease, reviewing the side effects of treatment and emphasizing the importance of social support for the cancer patient.

Table 1
Neuropsychological Assessment Battery

Tasks administered	
Clinical Interview with child	Clinical Interview with parent(s)
Kinetic Family Drawing (KFD)	Sentence Completion Task
Human Figure Drawings	NEPSY: All Core Subtests
Barkley Developmental History	Medical History
WISC-III Subtests: Similarities, Arithmetic, Picture Completion, Block Design	Child Behavior Checklist (CBCL)
Components of School Visit	
Teacher Interview	Teacher Report Form (TRF)
Instructional Styles Inventory (ISI)	Classroom Observation
Review of school records	
Additional Tasks	
Additional subtests of NEPSY Expanded Battery to further assess a particular domain	
Kaufman Test of Educational Achievement (Brief Form) K-TEA Brief; Kaufman Achievement Battery for Children (K-ABC): achievement subtests; Woodcock-Johnson Tests of Cognitive Ability (WJ-R): selected subtests	
Apperception or Self-Concept Tests: The Piers- Harris Children's Self-Concept scale, Robert's Apperception Test (R.A.T.): selected cards; Children's Apperception Test (Human Figures) (C.A.T.-H): selected cards	

Table 2
Classroom Presentation - Questionnaire

QUESTIONS ON LEUKEMIA

Please answer the following questions as True or False:

Leukemia is a type of cancer in which too many white blood cells are produced.

True False

Leukemia is treated by chemotherapy, bone marrow transplant, and radiation.

True False

Children with leukemia never get better.

True False

Children with leukemia may lose their hair and may feel like vomiting after their treatments.

True False

I can get sick from being around someone who has leukemia.

True False

The reintegration plans for the case studies consisted of practical advice on how best to help the child reach his or her potential in the classroom. The reintegration plans included:

1. Information regarding the child's learning/educational needs based on the results of the neuropsychological and mental processing results. Specific areas of strength and deficit were included, and recommendations were made regarding the learning style of each particular child. If this style did not match the instructional style of the teacher, suggestions as how best to work with this child were included.
2. Information regarding the patient's psychosocial functioning based on the clinical impression obtained during the assessment as well as the school visit.
3. Information regarding additional factors affecting each participant that might need attention. This might include physical weakness, prior learning disabilities, parental separation or divorce, and sibling rivalry.
4. An assessment of the availability of appropriate services in the school that might be utilized by the child.
5. A list of recommended interventions. These might include family therapy, marital counseling for the child's parents, educational therapy for the child, or suggested extramural activities, such as karate or art lessons.

Multiple follow-ups were conducted on several occasions following the feedback sessions. The aims of these follow-ups were to assess the usefulness of each reintegration program, as well as the school's attitude towards complying with this plan. If it was noted that some of the recommendations were impractical with respect to their implementation, appropriate modifications were suggested. In addition, at the conclusion of this study, the teachers and parents were each given a 5-item questionnaire to assess their satisfaction with the program.

Results

Neuropsychological Deficits

The results of the current study reflected nonspecific attention problems, particularly with regards to auditory attention and planning difficulties. Some difficulty maintaining attention on auditory tasks and attending to directions presented in a verbal format was evident throughout the neuropsychological assessments. This was noted particularly on tasks involving complex and multi-step instructions. Pertinent test scores from the test protocol of each of the three cases are outlined in Tables 3, 4, and 5. Many areas of difficulty identified were unique to the individual cases studied and are noted in Tables 6,7 and 8, which present abbreviated school reintegration plans for the participants in the study. As indicated in these tables, each child displayed specific areas of strength and weakness. For example, Kim exhibited difficulty maintaining consistent and accurate performance over long periods of time, along with slight difficulty in visualizing spatial relationships, in judging direction and estimating distance, and in understanding the relative positions of objects in space. She tended to be careless, sacrificing accuracy for speed, and her reduced accuracy affected her performance. She also displayed an extremely high level of distraction during class discussions or when working in a group. In contrast, Peter was noted to exhibit clear difficulty with the motor control of a pencil, and to display poor fine-motor and visual-motor skills and an inconsistent speed of performance. He was easily distracted when working on simple or monotonous tasks, particularly when the information was presented in a verbal format. When working on auditory tasks, this child required frequent repetition of instruction, although he did better when the information was socially relevant or meaningful. Susie's abbreviated reintegration plan indicates that she displayed extreme difficulty solving novel problems and was unable to complete tasks without a great deal of encouragement and repetition of instruction by the examiner. She was inconsistent in her level of performance and displayed a degree of impulsivity. She lacked age appropriate arithmetic and reading skills, scoring in the Below Average range on arithmetic and reading/decoding tasks.

Table 3
Pertinent Psychometric Data for Kim

NEPSY

CORE DOMAIN SCORES

	Attn/-Executive	**Lang.**	**Sensori-Motor**	**Visuo-Spatial**	**Memory**
Standard Score	92	107	98	112	123
Percentile	30	68	45	79	94
90% Conf. Interval	85-102	98-114	90-107	102-118	112-128

ATTENTION AND EXECUTIVE FUNCTION DOMAIN

Subtests	**Scaled Score**	**Percentile**
Tower	9	37
Auditory Attention and Response Set	8	25
Visual Attention	10	50

LANGUAGE DOMAIN

Subtests	**Scaled Score**	**Percentile**
Phonological Processing	13	84
Speeded Naming	9	37
Comprehension of Instructions	12	75

SENSORIMOTOR DOMAIN

Subtests	**Scaled Score**	**Percentile**
Fingertip Tapping	12	75
Imitating Hand Positions	9	37
Visuomotor Precision	8	25

VISUOSPATIAL DOMAIN

Subtests	**Scaled Score**	**Percentile**
Design Copying	15	95
Arrows	9	37

MEMORY DOMAIN

Subtests	**Scaled Score**	**Percentile**
Memory for Faces	17	99
Memory for Names	11	63
Narrative Memory	12	75

Table 4
Pertinent Psychometric Data for Peter

NEPSY

CORE DOMAIN SCORES

	Attn/-Executive	Lang.	Sensori-Motor	Visuo-Spatial	Memory
Standard Score	98	105	107	98	111
Percentile	45	63	68	45	77
90% Conf. Interval	90-107	97-112	97-114	90-107	102-117

ATTENTION AND EXECUTIVE FUNCTION DOMAIN

Subtests	Scaled Score	Percentile
Tower	12	75
Auditory Attention and Response Set	10	50
Visual Attention	7	16

LANGUAGE DOMAIN

Subtests	Scaled Score	Percentile
Phonological Processing	8	25
Speeded Naming	13	84
Comprehension of Instructions	12	75

SENSORIMOTOR DOMAIN

Subtests	Scaled Score	Percentile
Fingertip Tapping	12	75
Imitating Hand Positions	12	75
Visuomotor Precision	9	37

VISUOSPATIAL DOMAIN

Subtests	Scaled Score	Percentile
Design Copying	10	50
Arrows	9	37

MEMORY DOMAIN

Subtests	Scaled Score	Percentile
Memory for Faces	13	84
Memory for Names	15	9
Narrative Memory	7	16

Table 5
Pertinent Psychometric Data for Susie

NEPSY

CORE DOMAIN SCORES

	Attn/-Executive	Lang.	Sensori-Motor	Visuo-Spatial	Memory
Standard Score	92	95	87	92	105
Percentile 30	37	19	30	63	
90% Confid Interval	85-102	88-103	81-98	85-102	97-112

ATTENTION AND EXECUTIVE FUNCTION DOMAIN

Subtests	Scaled Score	Percentile
Tower	6	9
Auditory Attention and Response Set	10	50
Visual Attention	11	63

LANGUAGE DOMAIN

Subtests	Scaled Score	Percentile
Phonological Processing	7	16
Speeded Naming	8	25
Comprehension of Instructions	13	84

SENSORIMOTOR DOMAIN

Subtests	Scaled Score	Percentile
Fingertip Tapping	3	1
Imitating Hand Positions	7	16
Visuomotor Precision	15	95

VISUOSPATIAL DOMAIN

Subtests	Scaled Score	Percentile
Design Copying	9	37
Arrows	8	25

MEMORY DOMAIN

Subtests	Scaled Score	Percentile
Memory for Faces	13	84
Memory for Names	9	37
Narrative Memory	10	50

K-ABC

ACHIEVEMENT SUBTESTS SCALED SCORES

Achievement Subtest	Standard Score	Percentile Rank
Faces and Places	99 ± 1347	
Arithmetic 88 ± 11	21	
Riddles	94 ± 11	34
Reading/Decoding	81 ± 5	10

Table 6
Abbreviated School Reintegration Plan for Kim

Your child's cognitive, social, and emotional functioning was assessed by various means. The results of this assessment indicate that your child has strengths in the following areas: *attention to visual details; visuomotor integration; memory for faces.*

She has more difficulty in some other areas. These difficulties may have existed prior to the diagnosis of ALL or may be a direct result of the illness and its treatments. These difficulties are in the following areas: *auditory attention and memory; consistent and accurate performance over time; visualizing spatial relationships and judging direction; organization; focusing during discussions or lectures.*

Other factors that may be contributing to your child's difficulties include: *several life changes over the past two years; concerns about the health of her uncle and father.*

As a result of these difficulties, the following recommendations are made:

Specific classroom modifications, i.e. *supplementing verbal instructions with written direction; agreeing upon a signal between Kim and her teacher to help her refocus attention; presenting visual stimuli (pictures, charts, videotapes) along with auditory information.*

Educational therapy to improve study skills: *may be necessary if Kim continues to have difficulty organizing herself.*

Other: *educate Kim about her dad's illness; support group for parents; consistent and structured environment in her home and father's home; parents and teacher to read up about strategies to use with children with attention difficulties.*

Table 7
Abbreviated School Reintegration Plan for Peter

Your child's cognitive, social, and emotional functioning was assessed by various means. The results of this assessment indicate that your child has strengths in the following areas: *nonverbal reasoning; novel problem solving; remembering names; comprehension of simple verbal instructions.*

He has more difficulty in some other areas. These difficulties may have existed prior to the diagnosis of ALL or may be a direct result of the illness and its treatments. These difficulties are in the following areas: *focusing for long periods of time on tasks presented in a verbal format; fine-motor and visual-motor skills.*

Other factors that may be contributing to your child's difficulties include: *concerns about his darker complexion.*

As a result of these difficulties, the following recommendations are made:

Specific classroom modifications, i.e. *supplementing oral direction with written directions; active learning experiences; ongoing positive reinforcement for desired behaviors; frequent repetition of information; regular parent-teacher conferences; memory and reasoning games.*

Family therapy – *if conflict with siblings is ongoing.*

Other: *encourage Peter to complete his homework in a quiet room or library at a desk. Match Peter with a teacher next year whose style of teaching is similar to his current teacher's style.*

Table 8
Abbreviated School Reintegration Plan for Susie

Your child's cognitive, social, and emotional functioning was assessed by various means. The results of this assessment indicate that your child has strengths in the following areas: *nonverbal reasoning; understanding simple instructions; motor control of a pencil; memory for faces.*

She has more difficulty in some other areas. These difficulties may have existed prior to the diagnosis of ALL or may be a direct result of the illness and its treatments. These difficulties are in the following areas: *age appropriate arithmetic and reading skills; consistency of performance; solving novel problems; following complex and multi-step instructions.*

Other factors that may be contributing to your child's difficulties include: *lack of exposure to a formal academic program.*

As a result of these difficulties, the following recommendations are made:

Specific classroom modifications, i.e. *multisensory learning experiences; consistent feedback from her teacher; socially relevant learning experiences; group activities; breaking down complex directions and providing both written and verbal instruction.*

Educational therapy in the area of language
Educational therapy in the area of math
Educational therapy to improve study skills

Other: *Susie's social and school functioning should be re-evaluated at a later stage, and appropriate interventions sought if necessary.*

It should be noted that attention may be a confounding variable in studies evaluating the neuropsychological abilities of ALL patients, and it may be difficult to distinguish the effects of attentional problems from other cognitive deficits. It is thus plausible that an inconsistent attention level may have led to other difficulties noted among the participants in this study. It should also be noted that for any child, a poor attention span may interfere with his or her test-taking ability (Barkley, 1990). Therefore the results of these neuropsychological assessments may have been impacted by the child's attention difficulties. In addition, due to lack of pre-test measures, it was not possible to assess which, if any, of the child's cognitive functions may been unaffected by the experience of cancer and associated treatments. While premorbid levels of functioning can be assessed for adults with head injuries, this is not possible for children who are continually developing and are constantly acquiring knowledge at school (A. S. Kaufman, personal communication, February 16, 1999). This further complicates the issue of assessment of neuropsychological deficits.

All the children in this study had home tutoring during the period that they were absent from school. The tutor interviews were useful in that they provided the researcher with information about how the tutors viewed these children. However, not all the tutors appeared to have an accurate measure of the academic abilities of these children, and it seems that their lack of standardized assessment of the children's skills contributed to this inaccuracy. For example, although Susie's tutor indicated that her skills were at grade level, the results of the neuropsychological assessment clearly indicated that her academic achievement was very delayed. As a result of the tutor's inaccurate assessment, the child's parents were unaware of her poor academic skills and were surprised by the results of this assessment which reflected academic deficits. It would therefore be expected that without an accurate measure of a child's functioning prior to returning to school, the ALL patient may be faced with unreasonable expectations regarding his or her classroom functioning, possibly hindering school reintegration.

Psychosocial Deficits

One of the primary psychosocial difficulties noted in this study relates to the changes in the patient's view of him or herself as a result of experiencing a chronic illness. The life-threatening nature of childhood cancer, the chronicity of the treatment, and the possible long-term effects of treatment all generate a tangible handicap for the patient. However, a potential real handicap is the child's self-image, and the reaction of society to the child. Consistently, the children who participated in this study expressed their distress surrounding their hair loss, along with associated fears of social ridicule. These fears are expected and have been noted by other researchers in this area (Katz & Varni, 1993; Ross & Ross, 1984). During early childhood, a child tends to define him or herself in physical terms (Mussen, Conger, Kagan, & Huston, 1984). Descriptions of self involve observable physical features, such as hair color, and inner psychological experiences are not described as being separate from physical characteristics. Therefore, it would be expected that body image concerns might impact the child's self concept, even during these early childhood years. While the hair loss of the children in this study may have contributed to body image concerns, this does not seem to pose a threat to the emotional stability of the children. Despite their concerns, the participants appeared to maintain a strong sense of self-worth, as evidenced on the drawing, sentence completion and apperception tasks. This may be understood in terms of research into the content and structure of self-concept. This research has indicated that there are two components that underlie the development and maintenance of self-concept and self-worth (Harter, 1986). One of these components relates to the degree to which one is successful in those domains that one considers important, and the other component is based on the reflected appraisals of others, particularly parents and peers. Harter reported that children who maintain a positive sense of self-worth tend to discount the importance of domains in which they are not performing competently and endorse the importance of domains in which they are competent. In the current study, despite the stressors that they were experiencing, the children appeared to recognize and take pride in the manner in which they were able to cope with their illness, and at some point all proudly reported on how brave they were when being "stuck with the needle." They therefore seemed to focus on their own courage, for which they may have been praised by their parents and medical staff. This may reflect a tendency to slightly exaggerate their competence in an area, while discounting other areas of difficulty, such as changes in physical appearance. It appears that this is an adaptive strategy used by the children to maintain a high self-worth. In addition, it appears that appropriate social support from friends, teachers, and parents mediated these difficulties, which can be further understood in terms of a child's ability to maintain a strong self-worth by focusing on the appraisals of significant others.

One issue that affected the adjustment of the children in this study was the year long school absence after diagnosis, followed by frequent short absences during the maintenance phase of treatment. The participants reported feeling isolated from peers during their long absences; however, all made a remarkable adjustment back into their peer groups and seem to have been completely accepted upon their return to school. The frequent shorter absences appear to have created more academic rather than psychosocial difficulties for the children. This was particularly noted for Susie, who missed school on a frequent basis due to constant ill health and necessary treatments.

It is difficult to determine the extent to which any neuropsychological or psychosocial difficulty resulted from the experience of cancer and its treatments, or as a normal consequence of environmental stressors that are unrelated to the illness. Kim provided a good example of a child experiencing a range of psychosocial stressors aside from her ill health. This child experienced the divorce of her parents, her mother's subsequent remarriage, her father's chronic pancreatitis, and the diagnosis of a serious brain tumor in her uncle at the same time as the onset and initial treatment of her illness. Her neuropsychological difficulties, such as inattention, may therefore be only minimally related to her own health concerns.

Classroom Presentations

The aims of the classroom presentation were to educate the patient's classmates about his or her illness and to provide the students with opportunities to ask questions about the illness. Another goal of the presentation was to establish the child as the expert on his or her illness, while garnering support from his or her peers. There were a variety of factors that impacted the extent to which these goals were achieved. These include the children's level of cognitive development, their language skills, their preexisting knowledge about ALL, as well as their age. The classroom questionnaire (see Table 2) that was administered before and after the presentation allowed the examiner to assess the effectiveness of the presentation. The results of this questionnaire indicated that the presentations were clearly successful in educating the students about ALL. The post-test results reflected a 13% to 20% increase in correct responses, with the younger children displaying the highest increase in results. Those children in a more highly academic classroom environment initially obtained notably higher scores on the pretest thus showed a smaller increase in correct responses following the presentation. It appears that the less the children initially knew about ALL prior to the presentation and the better developed their language skills, the more effective the presentation was in increasing their knowledge in this area. It is important to remember, however, that in evaluating the effectiveness of these presentations it would be negligent to focus purely on the difference between pre- and post-test measures. One must consider that the presentations stimulated the thinking of the students, as indicated by their questions, and that this might lead to more long term changes in thinking about chronic illness and empathy towards chronically ill classmates.

One of the goals of the classroom presentation was that the patients would be recognized as experts of their illness. During the classroom presentation this important role was emphasized, with the hope that it would facilitate the child's acceptance by his or her peers. The child's individual status in the class appeared to be elevated by placing the child in a position of knowledge and power. Although the subjects in this study were at times hesitant about participating in these presentations, rather than feeling self-conscious, they appeared to thrive in a leadership role and take pride in their experience and their ability to transmit their knowledge to others. It would be expected that this would give the ALL patients approval and distinction among their classmates and help them maintain stronger self-esteem during this stressful period of their young lives.

Assessment of Instructional Style

Research into the interaction between aptitude and treatment suggests that the match between learner and instructor preferences influences student performance (Dowaliby, Curwin, & Quinsland, 1984; McInerney, McInerney, & Marsh, 1997). Aptitude-treatment-interaction (ATI) is based on the premise that there is no one best educational treatment or environment suited to a general, average individual, but that different individuals thrive in different environments suited to their own characteristics and needs. Numerous studies of ATI show that individuals learn more easily from one method than from another, and that differences between treatments are related to learner characteristics (Koran & Koran, 1984). The results of the ISIs administered were communicated to the teachers, who were provided with direct and practical feedback about the child's anticipated performance in response to his or her teaching style, along with suggested modifications to this style that would assist the child in reaching his or her academic potential. All teachers appeared interested and enthusiastic to learn about their own styles and were willing to consider appropriate changes to this style if necessary.

Parent/Teacher Evaluation of School Reintegration Programs

At the end of this study, each parent and teacher was asked to complete a questionnaire evaluating their perception of the effectiveness of the school reintegration programs. For all three subjects who participated in the study, the teachers and parents were open to the modifications suggested in the plans for school reintegration and reported that they found the reintegration programs to be of great benefit in evaluating and meeting the needs of these children. There were some differences in the parents' and teachers' appraisals of the value of the various components of the study. The component noted to be of most use by Kim's parents was the school observation since they reported obtaining clarity on how her behaviors impacted her classroom performance. Kim's teacher was excited by the information presented and particularly eager to follow through on the recommendations, noting that the recommendations would be useful for several of the other children in her class. On follow-up it appeared to the researcher that both Kim's parents and teacher would need to work at consistently implementing these modifications over a longer period of time in order to note a change in the child's behavior.

Peter's parents noted that they were very satisfied with the testing conducted and were able to implement many of the recommendations. They reported that the most useful part of the assessment was that they developed a better understanding regarding the needs of their son and the pace at which he works. Peter's teacher reported that she found the information yielded by the neuropsychological assessment of great use in that it highlighted the child's strengths and weaknesses, of which she had previously been unaware. Similar to Kim's teacher, Peter's teacher noted that many of the recommendations made by the researcher would be useful for the class as a whole. She further reported that she gained special insight from the teaching style inventory and noted that one of the most useful parts of the reintegration program was the classroom presentation which she reported as being very well received by her students.

Susie's parents reported that although they strongly agreed with the modifications recommended by the researcher as a result of the neuropsychological assessment, due to their child's continuing ill health and repeated hospitalizations they and the school had not yet had the opportunity to implement all the suggested recommendations and modifications. These parents noted that the most valuable part of the assessment was their child's relationship with the examiner, noting that Susie had formed a strong bond with the examiner that had helped in her school adjustment. Susie's teacher was eager to implement the suggested classroom modifications, but similarly to Susie's parents, she expressed concerns about this child's continued school absences due to her ongoing ill health. The teacher noted that in her opinion the primary issue was that Susie was comfortable at school, and that the importance of developing her academic skills would be stressed only once her health had improved.

Discussion

The design of this study was such that rather than focusing on group scores, as has been the focus of much prior research on ALL, the researcher was able to examine the individual needs of three ALL pediatric patients. This is compatible with Allport's (1961) idiographic approach, which emphasizes the value of the intensive study of the individual as a supplement to the study of groups. The study of the individual with ALL provides the researcher with detailed information not accessible from large-scale group studies. While the results of these single case studies cannot be generalized to the ALL population as a whole, these results provide a unique source of data that complements and contributes to other experimental research.

By limiting this study to three children, it allowed for the planning of specific interventions geared to the needs of the participants, which would not have been possible if group data had been used. The three children in this study clearly experienced a disruption in their schooling. This disruption has been noted in prior research assessing the academic needs of ALL patients, and Peckham et al. (1988) reported that as compared to younger children, children already in school when diagnosed with ALL experience more academic difficulties as a result of having their schooling interrupted.

The design of this study therefore provided the researcher with the opportunity of examining the integrity of the different components of neuropsychological functioning of these school-age children. This contributed to the understanding of the cognitive needs of the children. Rather than merely obtaining an IQ score as a measure of the child's ability, this study focused on obtaining a more in-depth and comprehensive perspective of each child's cognitive functioning, the results of which could then be presented to parents and teachers to assist them in understanding the needs of the child.

One of the strengths of the study is that data from multiple sources were integrated to create a clear understanding of the needs of the ALL patients. This involved formal and informal testing, observations in the home and school environment, and an assessment of the match between the child and his or her learning environment. A multiple assessment approach involves four important assessment procedures: norm-referenced tests, interviews, observations, and informal testing (Sattler, 1992). While normative data allows the researcher to determine significant cognitive, affective, and behavioral deviations, such data need to be supplemented with observations of referred children in their natural surroundings, the examiner's own clinical judgment about the child's behavior as a result of informal assessment, as well as interviews with the child, parents, teachers, and other individuals familiar with the child. Sattler notes that these four pillars of assessment complement one another and form a firm foundation for making decisions about children. In addition to utilizing a multiple assessment approach, in this study multiple informants were used when gathering information about the children. It is widely accepted that multi-informant assessment is more valid than relying on a single informant (Bell-Dolan & Allan, 1998). Power et al. (1998) demonstrate the incremental utility of combining informant reports in making diagnostic decisions about children, and quantify the degree to which diagnostic prediction is improved by combining the reports of parent and teacher reports. In the current study, the reports of teachers and medical staff, along with tutor interviews, complemented the reports of the parents, and helped the examiner view the child from a variety of perspectives.

The utility of multiple component programs to address the school reentry process of children with cancer has been demonstrated in prior research, for example Katz et al. (1988). Whereas the school reintegration programs developed in this study have several components in common with the School Reintegration Project developed by Katz et al. (1988), the current study included an evaluation of the patients' cognitive functioning. As noted above, this component provided important data that allowed the researcher to develop individualized reintegration programs that matched the needs of each patient, rather than a generic program for all ALL patients.

The patients in the current study were assessed during the maintenance phase of their treatments, within the first 3 years after diagnosis. It would be interesting to reassess these children at a later date to gather information about their acquisition of cognitive skills. The research into the neuropsychological functioning of pediatric cancer patients has reported an effect of elapsed time on the cognitive functioning of these patients and has further predicted a slowing down in the rate of acquisition of new skills (Mulhern, Wasserman, Fairclough, & Ochs, 1988; Radcliffe, Bunin, Sutton, Goldwein, & Phillips, 1994). It would therefore not be unexpected to note declines in cognitive functioning at follow-up periods as compared to their current functioning. Behaviors that may not presently be noted as a significant problem may subsequently lead to other problems, such as difficulty following directions, getting started on assignments, staying on task, asking for help, and completing tasks. Therefore, even if no learning disability is noted among children with ALL during the maintenance phase of their treatment, over time these children may begin to display some cognitive and performance difficulties which may lead to later problems with learning. This may not be unique to ALL patients but may also be evident among pediatric patients with other chronic illnesses who receive intensive CNS treatments.

Masera et al. (1997) discuss the importance of offering every child with leukemia a comprehensive program integrating medical, psychological, social, and educational components to ensure the highest possible quality of life.

Therefore, in order to meet the needs of the child with ALL, it appears that it is essential to provide each child and family with one psychosocial provider, such as a psychologist, who is responsible for providing a broad range of supportive mental health services. This professional should have special training in conducting comprehensive interviews with families, facilitating liaisons with other health care professionals, and understanding the impact of severe childhood chronic illness on the family. He or she should have neuropsychological assessment skills, competence with psychotherapy, and knowledge of childhood cancer. The primary role of this professional should be to help the child and his or her family adapt and cope with the illness and to assist with school reentry.

According to Deasy-Spinetta (1993), the future challenge is to make school intervention efforts an integral part of psychosocial care in all pediatric hematology/oncology centers. School problems of children with cancer can best be seen as part of the general problem of effective schooling for children with chronic illnesses, since the experiences and needs of children with cancer are similar to those of children with chronic disease (Larcombe et al., 1990). It is important to integrate the school, child, and parenting programs into the array of comprehensive health services offered to children with cancer. From among this array of services, individualized care plans can then be formulated to match specific child and family needs. The hope is that the components of the initial school reintegration program developed in this current study will be utilized and expanded upon in future research by psychologists as they strive to meet this challenge noted above.

Limitations of the Current Study and Implications for Future Research

In this study it was impossible to determine the extent to which any cognitive difficulty was the result of the experience of ALL and the associated treatments or if it had existed prior to the child's illness. The children in this study were in their early years of formal schooling, and none had had any prior assessments of their cognitive or academic skills. Since it is not possible to predict which children will develop ALL and to perform assessments on these children prior to their illness, future research in this area is limited. Although researchers have conducted assessments of patients shortly after diagnosis, and compared these results to those attained at follow-up (e.g., Copeland et al., 1988), it is impossible to determine the impact of disease-related factors at diagnosis on test scores. It may be useful to perform a similar study with older children who have had ongoing assessments in the school environment and to compare these results to assessments conducted after prophylactic treatment. It is likely, however, that these pre-illness assessments will involve group testing, such as the Stanford 9 or the CTBS. Both Anastasi (1988) and Aiken (1996) are critical of group testing, stating that it is more difficult to establish rapport and to monitor fatigue, cooperation, motivation, and the emotional state of the child in a group setting. Anastasi also notes that the examiner is not able to analyze the errors of the child during a group test. It therefore appears that group assessment may not be as effective as an individual battery in providing a clear picture of the child's potential or current functioning. In addition, group assessments at school primarily address academic skills and do not yield a profile of the child's neuropsychological performance. It therefore appears that it is commonly not possible to attain adequate pre-illness measures to examine the extent to which cognitive difficulty results from a chronic illness and the associated treatments. This problem is noted by Brown and Madan-Swain (1993), who indicate that a methodological problem of several studies on children with leukemia is a failure to report premorbid psychometric data.

An issue worth considering is that although the children did appear to make an excellent adjustment into the school environment, it is possible that due to the ill health of these children, their teachers may have unclear or lower expectations of their performance which may have biased teacher reports. It is possible that as the children move out of the maintenance phase of their treatment and continue to progress through school, more cognitive and behavior deficits may be reported, particularly as teachers' expectations of these children change over time. Deasy-Spinetta and Spinetta (1980) emphasize the need for frequent information exchanges between health care providers and teachers. Teachers need to know and be updated on the child's medical status, and Masera et al. (1995) note that teachers should be assured that their role is to teach and the hospital's role is to treat. This frees the teachers from unnecessary worries over the child's physical problems. The teacher plays a crucial role in facilitating the ill child's emotional and socio-academic adjustment in school by influencing the tone of the classroom and helping peers understand physical changes, absences, and potential limitations (Katz, 1980). It would therefore be important that teachers have the opportunity to discuss the child's illness with a health care professional on an ongoing basis, to express their feelings and obtain additional information. This will allow the teacher to set appropriate standards for the student's performance and to feel comfortable in enforcing those standards. It would also be useful to have regular meetings between the teacher, parent, and researcher to consider the changing needs of the student as he or she progresses through school and is faced with different challenges.

Conclusion

From the outset, the initial goal of this study was to conduct comprehensive psychoeducational assessments which would provide the researcher with substantial information regarding the ALL patient's cognitive and psychosocial functioning and would further guide the creation of appropriate reintegration programs. This goal was clearly attained during the study. In addition, however, further benefits were also achieved beyond this initial objective. The entire process of interacting with the child and his or her family and designing reintegration plans affected the child in a manner that was unanticipated at the outset of the study. The child was provided with the support and guidance of the researcher, and this served as a growth enhancing experience for the child. In addition, the family benefitted from this project in that it had a therapeutic effect on the entire family system. While family members were provided with new information about the ill child, their own beliefs about the child were also validated, and parents were reassured that their observations about their child were well-founded. They were also assured of their abilities to care for their child and to hold the family together throughout this critical period. The study further served as a therapeutic intervention for the teacher, in that he or she was educated about his or her teaching style, was provided with information about how to cope with the chronically ill child, and was offered appropriate support and encouragement. As a result of this, the presence of the child was not experienced as an added burden by these teachers, and they were motivated to meet the needs of the child. The medical staff, who are not ordinarily able to track the child's school adjustment, were reassured that the psychosocial and cognitive needs of the child were being addressed, and they experienced pleasure in serving as a component of this life enhancing program for their patients. This study therefore facilitated the cooperation between a variety of professionals and health care providers working with the child. However, at the foundation was the comprehensive assessments of each child, without which these additional gains would not have been attained.

The entire process of assessing the child, working with parents and teachers, and planning the reintegration projects therefore had a cumulative effect. This was an unanticipated bonus of the study, which enhances the literature regarding therapeutic interventions for ALL children returning to their schools.

References

Achenbach, T.M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T.M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Aiken, L.R. (1996). *Assessment of intellectual functioning. Perspectives on individual differences* (2nd Ed.). New York, NY: Plenum Press.

Allport, G.W. (1961). *Pattern and growth in personality.* New York, NY: Holt, Rinehard & Winston.

Anastasi, A. (1988). *Psychological testing* (6th Ed.). New York, NY: MacMillan

Barkley, R.A. (1990). *Attention-deficit hyperactivity disorder.* New York, NY: Guilford Press.

Bell-Dolan, D.J., & Allan, W.D. (1998). Assessing elementary school children's social skills: Evaluation of the parent version of the Matson Evaluation of Social Skills with Youngsters. *Psychological Assessment, 10*(2), 140-148.

Brouwers, P., Riccardi, R., Poplack, D.G., & Fedio, P. (1984) Attentional deficits in long-term survivors of childhood acute lymphoblastic leukemia (ALL). *Journal of Clinical Neuropsychology, 6*(3), 325-336.

Brown, R.T., & Madan-Swain, A. (1993). Cognitive, neuropsychological and academic sequelae in children with leukemia. *Journal of Learning Disabilities, 26*(2), 74-90.

Cairns, N. U., Klopovich, R.N., Hearne, E., & Lansky, S.B. (1982). School attendance of children with cancer. *Journal of School Health, 52*, 152-155.

Canfield, A.A., & Canfield, J.S. (1988). *Manual for the Instructional Styles Inventory (ISI).* Los Angeles, CA: Western Psychological Services.

Charlton, A., Larcombe, I.J., Meller, S.T., Morris Jones, P.H., Mott, M.G., Potton, M.W., Tranmer, M.D., & Walker, J.J.P. (1991). Absence from school related to cancer and other chronic conditions. *Archives of Disease in Childhood, 66*, 1217-1222.

Ciesielski, K.T., Yanofsky, R., Ludwig, R.N., Hill, D.E., Hart, B.L., Astur, R.S., & Snyder, T. (1994). Hypoplasia of the cerebellar vermis and cognitive deficits in survivors of childhood leukemia. *Archives of Neurology, 51*, 985-993.

Copeland, D.R., Dowell, R.E., Fletcher, J,M., Sullivan, M.P., Jaffe, N., Cangir, A., Frankel, L.S., & Judd, B.W. (1988). Neuropsychological test performance of pediatric cancer patients at diagnosis and one year later. *Journal of Pediatric Psychology, 13*(2), 183-196.

Cousens, P., Waters, B., Said, J., & Stevens, M. (1988). Cognitive effects of cranial radiation in leukemia: a survey and meta-analysis. *Journal of Child Psychology and Psychiatry, 29*, 839-852.

Deasy-Spinetta, P.M. (1993). School issues and the child with cancer. *Cancer, 71*, 3261-3264.

Deasy-Spinetta, P.M., & Spinetta, J.J. (1980). The child with cancer in school: Teacher's appraisal. *The American Journal of Pediatric Hematology-Oncology, 2*, 89-94.

Dowaliby, F.J., Curwin, R. & Quinsland, L. (1984). Classroom structure and student participation: An aptitude-by-treatment interaction approach to instructional research for the hearing-impaired. *American Educational Research Journal, 21*(4), 807-824.

Fletcher, J.M., & Copeland, D.R. (1988). Neurobehavioral effects of central nervous system prophylactic treatment of cancer in children. *Journal of Clinical and Experimental Neuropsychology, 10* (4), 495-538.

Gortmaker, S.L., Walker, D.K., Weitzman, M., & Sobol, A.M. (1990). Chronic conditions, socioeconomic risks, and behavioral problems in children and adolescents. *Pediatrics, 85*, 267-276.

Harter, S. (1986). Processes underlying the construct, maintenance, and enhancement of the self-concept in children. in J. Suls & A. Greenwald (Eds.), *Psychological Perspectives of the Self* (Vol. 3, pp. 137-181). Hillsdale, NJ: Erlbaum.

Katz, E.R. (1980). Illness impact and social reintegration. In: Kellerman, J. (Ed.), *Psychological aspects of childhood cancer*. (pp.14-46). Springfield, Illinois: Charles C. Thomas.

Katz, E.R, Rubinstein, C.L., Hubert, N.C., & Blew, A. (1988). School and social reintegration of children with cancer. *Journal of Psychosocial Oncology, 6*(3/4), 123-139.

Katz, E.R., & Varni, J.W. (1993). Social support and social cognitive problem-solving in children with newly diagnosed cancer. *Cancer, 71*, 3314-3319.

Katz, E.R., Varni, J.W., Rubenstein, C.L., Blew, A., & Hubert, N. (1992). Teacher, parent, and child evaluative ratings of a school reintegration intervention for children with newly diagnosed cancer. *Children's Health Care, 21*(2), 69-75.

Kaufman, A.S., Kaufman, J.C., Balgopal, R., & McLean, J.E. (1996). Comparison of three WISC-III short forms: Weighing psychometric, clinical, and practical factors. *Journal of Clinical Child Psychology, 25*(1), 97-105.

Koran, M.L., & Koran, J.J. (1984). Aptitude-treatment-interaction in science education. *Journal of Research in Science Teaching, 21*(8), 793-808.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment*. San Antonio, TX: The Psychological Corporation.

Lansky, S.B., Cairns, N.U., & Zwartjes, W. (1983). School Attendance among children with cancer: A report from two centers. *Journal of Psychosocial Oncology, 1*(2), 75-83.

Larcombe, I.J., Walker, J., Charlton, A., Meller, S., Morris Jones, P., & Mott, M.G. (1990). Impact of childhood cancer on return to normal schooling. *BMJ, 301*, 169-171.

Madan-Swain, A., & Brown, R.T. (1991). Cognitive and psychosocial sequelae for children with acute lymphoblastic leukemia and their families. *Clinical Psychology Review, 11*, 267-294.

Masera, G., Jankovic, M., Adamoli, L., Corbetta, A., Fraschini, D., Lia, R., Collino, L., Locati, A., Pertici,S., Bissi, R., Bertolini, M., Verga, G., & Spinetta, J.J. (1997). The psychosocial program for childhood leukemia in Monza, Italy. In F.F. Holmes & J.J. Kepes (Eds.). Challenges and Opportunities in Pediatric Oncology. *Annals of the New York Academy of Sciences, 824* (pp. 210-220). New York, NY: New York Academy of Sciences.

Masera, G., Jankovic, M., Deasy-Spinetta, P., Adamoli, L., Ben Arush, M.W., Challinor, J., Chesler, M., Colegrove, R., Van Dongen-Melman, J., McDowell, H., Eden, T., Epelman, C., Kingma, A., Morris Jones, P., Nesbit, M.E., Reynolds, H., Schuler, D., Stevens, M., Vasankari-Vayrynen, L., Wilbur, J.R., & Spinetta, J.J. (1995). SIOP working committee on psychological issues in pediatric oncology: Guidelines for school/education. *Medical and Pediatric Oncology, 25*, 429-430.

McInerney, V., McInerney, D.M., & Marsh, H.W. (1997). Effects of metacognitive strategy training within a cooperative group learning context on computer achievement and anxiety: An aptitude treatment interaction study. *Journal of Educational Psychology, 89*(4), 686-695.

Meadows, A.T., Massari, D., Fergusson, J., Gordon, J., Littman, P., & Moss, K.(1981). Declines in IQ scores and cognitive dysfunctions in children with acute lymphocytic leukemia treated with cranial irradiation. *Lancet, 7*, 1015-1018.

Moore, I.M., Kramer, J.H., Wara, W., Halberg, F., & Ablin, A.R. (1991). Cognitive function in children with leukemia. *Cancer, 68*, 1913-1917.

Mulhern, R.K., Ochs, J., & Fairclough, D. (1992). Deterioration of intellect among children surviving leukemia: IQ test changes modify estimates of treatment toxicity. *Journal of Consulting and Clinical Psychology, 60* (3), 477-480.

Mulhern, R.K., Wasserman, A.L., Fairclough, D., & Ochs, J. (1988). Memory function in disease-free survivors of childhood acute lymphocytic leukemia given CNS prophylaxis with or without 1,800 cGy cranial irradiation. *Journal of Clinical Oncology, 6*(2), 315-320.

Mussen, P.H., Conger, J.J., Kagan, J., & Huston, A.C. (1984). *Child development and personality* (6th Edition). New York, NY: Harper & Row.

Noll, R.B., Gartstein, M.A., Vannatta, K., Correll, J., Bukowski, W.M., & Davies, W.H. (1999). Social, emotional, and behavioral functioning of children with cancer. *Pediatrics, 103*(1), 71-78.

Noll, R.B., Bukowski, W.M., Rogosh, F.A., LeRoy, S., & Kulkarni, R. (1990). Social interactions between children with cancer and their peers: Teacher ratings. *Journal of Pediatric Psychology, 15* (1), 43-56.

Peckham, V.C., Meadows, A,T.,Bartel, N., & Marrero, O. (1988). Educational late effects in long-term survivors of childhood acute lymphocytic leukemia. *Pediatrics, 81*(1), 127-133.

Poplack, D.G. (1989). Acute Lymphoblastic Leukemia. In P.A. Pizzo, & D.G. Poplack (Eds.). *Principles and Practices of Pediatric Oncology.* (pp. 323-366).

Power, T.J., Andrews, T.J., Eiraldi, R.B., Doherty, B.J., Ikeda, M.J., DuPaul, G.J., Landau, S. (1998). Evaluating attention deficit hyperactivity disorder using multiple informants: The incremental utility of combining teacher with parent reports. *Psychological Assessment, 10*(3), 250-260.

Radcliffe, J.B., Bunin, G.R., Sutton, L.N., Goldwein, J.W., & Phillips, P.C. (1994). Cognitive deficits in long-term survivors of childhood medulloblastoma and other noncortical tumors: age - dependent effects of whole brain radiation. *International Journal of Developmental Neuroscience, 12*(4), 327-334.

Reynolds, C.R., & Kamphaus, R.W. (1992). *Manual for the Behavior Assessment System for Children.* Circle Pines, MN: American Guidance Service.

Ross, D.M., & Ross, S.A. (1984). Teaching the child with leukemia to cope with teasing. *Issues in Comprehensive Pediatric Nursing, 7*, 59-66.

Rynard, D. W., Chambers, A., Klinck, A.M., & Gray, J.D. (1998). School support programs for chronically ill children: Evaluating the adjustment of children with cancer at school. *Children's Health Care, 27*(1), 31-46.

Sattler, J. M. (1992). *Assessment of Children* (3rd Edition). San Diego, CA: Jerome M. Sattler.

Soni, S.S., Marten, G. W., Pitner, S.E., Duenas, D.A., & Powazek, M. (1975) Effects of central-nervous-system irradiation of neuropsychologic functioning of children with acute lymphocytic leukemia. *The New England Journal of Medicine, 293*(3), 113-118.

Stehbens, J.A., Kaleita, T.A., Noll, R.B., MacLean, W.E., O'Brian, R.T., Waskerwitz, M.J., & Hammond, G.D. (1991) CNS prophylaxis of childhood leukemia: What are the long-term neurological, neuropsychological, and behavioral effects? *Neuropsychology Review, 2*(2), 147-177.

Stehbens, J.A., Kisker, C.T., & Wilson, B.K. (1983). School behavior and attendance during the first year of treatment for childhood cancer. *Psychology in the Schools, 20*, 223-228.

Taylor, H.G., Albo, V.C., Phebus, C.K., Sachs, B.R., & Bierl, P.G. (1987). Postirradiation treatment outcomes for children with acute lymphocytic leukemia: Clarification of risks. *Journal of Pediatric Psychology, 12*(3), 395-411.

Thompson, R.J., & Gustafson, K.E. (1995). *Adaptation to chronic childhood illness.* Washington, DC: American Psychological Association.

Varni, J.W. (1993). *Perceived physical appearance in children and adolescents with limb deficiencies.* Keynote address at the University of New Brunswick's International Symposium on Myoelectic Control: Future Trends in Myoelective Technology, New Brunswick, Canada.

Varni, J.W., Katz, E.R., Colegrove, R., & Dolgin, M. (1995). Perceived physical appearance and adjustment of children with newly diagnosed cancer: A path analytic model. *Journal of Behavioral Medicine, 18*(3), 261-278.

Wechsler, D. (1991). *Manual for The Wechsler Intelligence Scale for Children – Third Edition.* San Antonio, TX: The Psychological Corporation.

Whitt, K.J., Wells, R.J., Lauria, M.M., Wilhelm, C.L., & McMillan, C.W. (1984). Cranial radiation in childhood acute lymphocytic leukemia. *American Journal of Diseases of Children, 138*, 730-736.

Williams, J.M., & Davis, K.S. (1986). Central nervous system prophylactic treatment to childhood leukemia: Neuropsychological outcome studies. *Cancer Treatment Reviews, 13*, 1-15.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 1, 45-50

Correlates of School Academic Success: A State Report Card Study

Jerry G. Mathews
Auburn University

Data from a state report card were used to determine which variables characterized bivariate and multivariate relationships between school quality indicator variables and academic status in high and low SES schools and high and low achievement schools. A grouping variable, academic status, and 15 other variables published in the report card were used in the analysis. Thirteen hundred nineteen schools in 127 school districts were available in the data which were extracted from a world wide web page and represented all the schools in the state. Discriminant analysis and Fisher's z_r transformations indicated that SES variables moderate the relationships between five indicator variables that were correlated with academic status. The study provided information using variables in the state report card that may address the needs of at-risk students, improve student achievement, and guide future policy for improving school quality.

The 1960s and 1970s were decades of educational reform that introduced standard setting in accountability movements (Porter, 1992). Porter noted that the National Council on Educational Standards and Testing (NCEST) recommended national educational standards to raise the expectations that teachers have for students and to hold students accountable for academic progress. These efforts to establish standards brought on a new enthusiasm for assessments of student achievement and school accountability (Children First Act of 1988; Connecticut State Department of Education, 1993; Maryland State Department of Education, 1989; Mississippi State Department of Education, 1993). Kirst (1990) reported the following about the origins of accountability:

> Accountability has roots in many areas of management, including theories about incentives and business concepts. Before educators borrowed the term and imbued it with their own additional meanings, accountability expressed a relationship between those who controlled institutions and those who had power to displace them. (p. 5)

Brown (1990) identified three traditional concerns in educational accountability that emerged in the 1980s. These three concerns were (a) what children were taught, (b) the cost of education, and (c) employment of qualified personnel.

The Maryland State Department of Education (1989) defined accountability for education as "attempts to measure, disclose, interpret, and improve results achieved by public school education programs" (p. 3). The Maryland State Department of Education implemented an accountability program mandated by a law passed by the Maryland state legislature in 1972 called the *Educational Accountability Act*. This act provided "for the establishment of a statewide accountability system to ensure that educational programs lead to attainment of established goals and objectives and to provide information for the analysis of differential effectiveness of instructional programs" (p. 3).

Since the initial legislation of the 1970s, accountability in public schools has been on the increase mainly in similar forms of state imposed accountability systems (Franklin & Crone, 1993a; Franklin & Crone, 1993b). The Council of Chief State School Officers (1996) reported that as of 1996, 52 of 54 state education agencies had at least one annual accountability or indicator report, while 32 states now have some form of policy legislation enacted that require the reporting of school district level or school building level performance indicators. Currently, many states report both district level and building level data in the format of a state report card (Bobbett, French, & Achilles, 1992a; Bobbett, French, & Achilles, 1992b; Bobbett, French, & Achilles, 1992c; Bobbett, French, & Achilles, 1993; Bobbett, French, & Achilles, 1994; Kochan, 1993; Mathews & Hackett, 1997; Mathews, Ross, & Spencer, 1998; Raivetz, 1992).

Jerry G. Mathews is an Assistant Professor in Educational Foundations, Leadership and Technology at Auburn University. Correspondence regarding the article should be sent to Jerry G. Mathews, Educational Foundations, Leadership and Technology, 4016 Haley Center, Auburn University, AL 36849-5221 or by e-mail at matheig@mail.auburn.edu.

A report card from a southern state containing school quality indicators was chosen as the topic of this study. The report card received wide-spread media attention when it was released to the public by the State Department of Education (SDE). The state's public newspapers reported that schools scoring at caution, alert, and failure on the Stanford Achievement Test, Ninth edition (SAT9) had received warnings from the SDE concerning academic accreditation status. In a newspaper release (Markley, G., 1997, November 14), the State Superintendent informed the public that "We've raised the bar . . . (at many levels) [sic] . . . . The report cards help us see where we are and where we need to be in meeting these higher standards" (p. A1). The news article further noted that "schools that score low [on the SAT9] for the second year in a row will draw 'immediate attention' from his [state superintendent's] office, and he expects to see 'dramatic improvement' within a reasonable time" (p. A2). The state superintendent added that school leaders [principals and superintendents] would be held accountable for their schools' progress. While accountability in schools is the responsibility of administrators, the state superintendent's statement falls short of the standards recommended by the National Council of Educational Standards and Testing; that is, administrators, teachers, and students must all be held accountable for academic progress. Will simply raising the bar in standardized testing result in higher academic achievement for schools that are not performing at desired levels to begin with? The superintendent's statements appear to establish sanctions for low achieving schools rather than providing plans for academic improvement.

A study by Young and Smith (1997) indicated that poor school outcomes have been linked to students from racial/ethnic minority backgrounds and low income families. The theme for this piece of literature highlighted three decades of similar studies that "have linked the educational disadvantage of minority students to a combination of out-of-school [contextual] factors, many of which center on family characteristics, such as poverty and parents' education" (p. 1). The report recommended that policy makers should be aware of differences in the background of students, as well as differences in the climate and resources of schools, in order to provide effective improvement measures.

Socioeconomic status (SES) has a powerful influence on student achievement and attainment. It would be an unfair disadvantage to some schools if SES were not taken into consideration. A school with the majority of its students in the high SES category is more likely to have higher SAT scores, lower drop-out rates, and higher performance in general areas than a school with the majority of its students in the low SES category. The teachers, principals, and parents are faced with different in-school and out-of-school factors in schools with students of different SES levels (Council for School Performance, 1999).

From a school quality standpoint, do report card data provide policy-relevant information for low and high achieving schools and for low and high socioeconomic schools? This study was conducted to determine what report card variables differentiate between low and high achieving schools and low and high SES schools. The study was conducted in two phases. The initial purpose of this study was to identify from an initial set of 15 indicator variables a reduced set of school quality indicator variables at the school building level that correlated with an educational outcome variable, academic status, in a state report card. The main purpose of the study was to investigate group differences in bivariate and multivariate relationships between the reduced set of indicator variables and academic status in high and low socioeconomic status schools.

Method

The data for this study were extracted from a state report card data file which was retrieved from an Internet world wide web site. The report card data included 16 major variables representing academic achievement, socioeconomic status, financial status and teacher and student variables from 1319 schools in 127 districts. Of these 16 variables, 15 were designated as school quality indicator variables and one as an outcome variable.

The 15 initial school quality indicator variables were (a) ACT test scores, (b) ability-achievement comparison index, (c) average daily attendance, (d) expenditures per pupil, (e) high school exit exam, (f) mill equivalent, (g) mill equivalent yield per mill, (h) percent local revenues, (i) percent federal revenues, (j) percent state revenues, (k) professional certification, (l) projected 4-year dropout, (m) Stanford Achievement Test total battery, (n) student enrollment and (o) writing tests.

The outcome variable, academic status, in this study was dichotomized and defined as follows: (a) *high achieving* schools were those receiving academic clear or academic exempt status and (b) *low achieving* schools were those receiving academic caution, alert or failure status. In this report card publication, the SDE assigned each school an academic status based on test results of the SAT9 total battery as percentile rank scores in Grades 3 through 11. In the report card, academic status had five categories: (a) exempt, (b) clear, (c) caution, (d) alert and (e) failure. Exempt and clear schools were defined as schools scoring at the 40th percentile rank or above on the SAT9 overall total battery. Caution, alert, and failure schools were defined as having an SAT9 percentile rank of 39 or below.

To receive a caution status, over 50% of the students in the school must score below the 40th percentile on the SAT9. A school will receive an Alert 1 status if over 50% of the student body scores below the 23rd percentile on the SAT9 or if a school that was on caution status the previous year fails to make the suggested improvements outlined by the SDE. The suggested improvement by the SDE is that more students pass the SAT9. A school will receive an Alert 2 status if over 50% of the student body scores below the 23rd percentile on the SAT9 for the second year or if a school that was on caution status has failed to make the suggested improvements for the second year. A school will receive an Alert 3 status if over 50% of the student body scores below the 23rd percentile on the SAT9 for the third year or if a school that was on caution status has failed to make suggested improvements for the third year. The SDE sanction is intervention at this point. [These grading definitions are documented in the state superintendent's report card for each fiscal year of publication. This document is available to the reader upon request. The source citation is excluded from this manuscript to preserve the anonymity of the state under study.]

Analysis and Results

The first phase of the analysis focused on identifying which of the original 15 indicator variables were related to the outcome variable, academic status. The 15 initially identified indicator variables were further reduced to10 indicator variables because five of these variables were grade level specific and were not available across all school levels. For analysis purposes, 319 schools were randomly sampled from the total population of 1319 schools in the data set. This 319-school data set served as an initial screening sample for further variable selection but was not used in later statistical analytical procedures. A limitation of this procedure is the result of possible spurious correlations. Even though bivariate normality appeared to be reasonable in this analysis based on scatterplots, linear associations may result because one of the variables may be correlated with a third variable. Hence, there is no automatic conclusion that correlation implies causation. Table 1 summarizes these bivariate correlations.

As can be seen in this table, 8 of the 10 indicator variables were found to be significantly related to the outcome variable, academic status. The percent state revenue and student enrollment indicator variables were eliminated from the list because they did not correlate with the criterion variable. Therefore, a set of eight continuous indicator (independent) variables remained from the initial set of 15 indicator variables.

The second phase of the study was to determine bivariate and multivariate relationships between the eight remaining indicator variables and academic status in high and low SES schools. First, the remaining data set of 1000 schools was divided into two socioeconomic categories: low SES and high SES. Schools falling below the 50th percentile of students eligible for free and reduced lunches (*n* = 462) were identified as high SES. Schools at or above the 50th percentile of students eligible for free or reduced lunches (*n* = 538) were identified as low SES.

Table 1
Bivariate Correlations Between School Quality Indicators and Academic Status

Variables	2	3	4	5	6	7	8	9	10	11
1. Academic status	.417*	.060	.223*	.214*	-.003	-.464*	.214*	.161*	-.140*	-.264*
2. Ability/achievement comparison index		-.294*	.396*	.121*	-.071	-.135*	.116*	.121*	.017	.142*
3. Student enrollment			-.030	.097	-.029	-.160*	-.018	-.020	-.137*	.081
4. Average daily attendance				.077	-.065	-.049	.135*	.125*	-.005	.166*
5. Percent local revenues					-.887*	-.601*	.583*	.229*	.520*	.056
6. Percent state revenues						.166*	-.561*	-.153*	-.719*	.013
7. Percent federal revenues							-.276*	-.219*	.137*	-.140*
8. Mill equivalent								.196*	.344*	.140*
9. Mill equivalent mill per yield ($)									.093	.145*
10. Expenditures per pupil										.016
11. Percent teachers with advanced degrees										

*$p < .05$

Within the low SES group, a discriminant analysis was conducted with the eight indicator variables and the dichotomously scored outcome measure, academic status. All eight indicator variables were entered into the equation simultaneously. The obtained statistically significant canonical correlation was .581, $p < .001$ which indicates a strong correlation between the indicator variables discriminant score and academic status in the low SES group.

Next, the equation produced within the low SES group was then applied to the high SES group. This procedure obtained weighted combinations of indicator variables that were then correlated with academic status in the high SES group. This cross-validated statistically significant correlation coefficient in the high SES group was .237, $p < .001$.

The canonical correlation coefficients for the low SES group and for the high SES group were then converted to Fisher's z_r values. The difference between the two Fisher's z_r values was computed, and this difference was divided by the standard error to produce a critical z. The obtained z-value of 4.327 indicated that the multivariate relationship between the eight indicator variables and academic status was significantly ($p < .001$) higher in the low SES group than in the high SES group.

A second discriminant analysis was conducted using the high SES group. All eight indicator variables were entered into the equation simultaneously. The equation produced within the high SES group was then applied to the low SES group. This procedure obtained weighted combinations of indicator variables that were then correlated with academic status in the high SES group. The obtained statistically significant canonical correlation in the high SES group was .274, $p < .001$. The equation was then applied to the low SES group and, again, the analysis correlated the weighted combination of the eight indicator variables with the dichotomously scored outcome measure, academic status. This cross-validated statistically significant correlation coefficient in the low SES group was .513, $p < .001$.

The canonical correlation coefficients for the low SES group and for the high SES group were then converted to Fisher's z_r values. The difference between the two Fisher's z_r values was computed, and this difference was divided by the standard error to produce a critical z. The obtained z-value of 4.341 indicated that the multivariate relationship between the eight indicator variables and academic status was significantly higher in low SES group, $p < .001$. Even when the equation was optimized for the high SES group, the weighted combination of the eight indicator variables for the low SES group correlated higher with academic status ($r = .513$) than the high SES group ($r = .274$). The procedure confirmed the result of the first discriminant analysis that academic status and the weighted combination of indicator variables were more highly correlated in the low SES schools than for the high SES schools.

In the final analytical procedure, bivariate correlations were used to test for significant relationships between the eight school quality indicator variables and the outcome variable, academic status, for low SES and high SES schools separately. The correlation coefficients between the eight indicator variables and academic status for the low SES group and the high SES group were converted to Fisher's z_r values. The difference between the two Fisher's z_r values was computed, and this difference was divided by the standard error to produce a critical z-value for each SES pairing of the eight variables. Table 2 shows the results of this procedure.

Table 2
Significance Tests of Correlations Between School Quality Indicators and Academic Status of Low SES and High SES Schools Using Fisher's z_r Transformations

	Low SES		High SES		
Variable	r	z_r	r	z_r	z
Ability/achievement comparison index	.461**	.479	.189**	.192	4.697***
Average daily attendance	.164**	.167	.054	.055	2.338***
Percent local revenues	.013	.013	-.090*	-.090	1.188
Percent federal revenues	-.184**	-.186	-.001	-.001	-2.855***
Mills equivalent	.084	.084	-.005	-.005	1.373
Mill equivalent yield per mill ($)	.166**	.168	-.092	-.082	3.858***
Percent teachers higher degree	.212**	.215	.093*	.093	2.000***
Expenditure per pupil	-.171**	-.173	-.151**	-.152	-0.324

* $p < .01$, ** $p < .05$, ***critical z-value

As indicated by this table, correlations between five indicator variables and academic status were significantly different in the low SES group and the high SES group. In all instances, the correlations were higher in the low SES group than in the high SES group.

Discussion

The results of this study have shown that SES is an important variable in moderating relationships between school quality indicator variables and academic status on the state report card. The indicator variables, ability/achievement comparison index, average daily attendance, mill equivalent yield per mill, percent of teachers with higher degrees and per pupil expenditure were significant correlates of school academic status. These five variables were substantially more important correlates of academic success for low SES schools than for the high SES schools. For the low SES, low achieving schools, these same indicators warrant the attention of education decision makers for this state.

According to the results of this study, the ability/achievement comparison index indicator variable has a stronger correlation with academic status in the low SES schools than in the high SES schools. The low SES schools may be achieving close to the expected ability but are still categorized in the caution and alert academic status. High SES schools could possibly have lower achievement scores than what is expected and still be categorized in clear academic status.

This study shows that student attendance is significantly more important in achieving higher academic status in low SES schools than in high SES schools. Absence for students in the low SES schools may have a more negative impact on achievement than in the high SES schools. It is plausible that opportunities may exist for students of high SES schools to be invited to attend educational enrichment activities when they are absent from school to a greater extent than is the case for students in low SES schools. For example, more affluent families may have their children accompany them to professional conferences or on business trips that provide the opportunity to visit museums, aquariums, historical sites or other educational enrichment activities. It is less likely that less affluent families are able to provide their children with educational enrichment opportunities related to absence from school. For example, the student's absence from school may result in merely watching television programs that provide no educational value.

Historically, a higher percentage of federal revenues are associated with low SES schools. Schools with students at, or near, the poverty level qualify for more federal funds that high SES schools. In this study, the relationship between higher federal revenues and academic status is more important in low SES schools than for high SES schools. Low SES schools have higher federal revenues yet lower academic status than the high SES schools as would be expected. Federal revenues account for a large portion of the revenues of low SES schools including funds for free and reduced lunches. It is possible that in high SES schools a greater degree of out of school learning takes place than in low SES schools.

The findings in this study are consistent with the call for equity funding and transferring more resources into low SES schools that lack sufficient funding. The mill equivalent yield per mill is significantly more important when correlated with academic status in low SES schools than in the high SES schools. It is reasonable to assume that low SES schools would benefit from higher millage revenues, attain higher achievement and, hence, a higher academic status on the state report card.

The indicator variable, percentage of teachers with higher degrees, provides more important information for low SES schools than for the high SES schools when the variable is correlated with academic status. Schools with a higher percentage of teachers with higher degrees are associated with higher academic status. It may be concluded that low SES schools with a higher percentage of teachers with advanced degrees may better prepare students academically and provide more effective learning opportunities.

The relationship between expenditure per pupil relationship and academic status is higher in low SES schools than in the high SES schools. The per pupil expenditure variable is a funding variable that is possibly useless for analysis because this variable is an aggregate variable of other important funding variables. The federal revenues variable may serve to inflate the total per pupil expenditure variable for low SES schools. A previous study (Mathews & Hackett, 1997) shows that the high SES schools have higher instructional expenditures than low SES schools even though the total per pupil expenditure may be even higher for the low achieving, low SES schools, possibly because of the inflation by federal funds.

This state report contained no published information as to how the contextual report card variables were utilized for improving school quality. However, multivariate analysis of contextual report card indicator variables tends to demonstrate that contextual variables provide important information in the relationship between these contextual variables and academic status in low SES schools. Analysis and study of contextual report card variables may provide SDE policy makers and planners a conceptual framework to improve student achievement. The utility of these contextual variables in assessing school quality is influenced by the socioeconomic status of the schools.

References

Bobbett, G. C., French, R. L., & Achilles, C. M.. (1992a, November). *Evaluation of the categories currently used in report cards with student outcome.* Paper presented at the annual meeting of the Southern Regional Council of Educational Administration, Atlanta, GA. (ERIC Document Reproduction Service No. ED 352 417)

Bobbett, G. C., French, R. L., & Achilles, C. M.. (1992b, February). *Student outcome and policy making: An analysis of Tennessee's report cards on schools (why do we bet on .25 when .75 is running loose)?* Paper presented at the annual meeting of the American Association of School Administrators, San Diego, CA.

(ERIC Document Reproduction Service No. ED 342 804)

Bobbett, G. C., French, R. L., & Achilles, C. M.. (1992c, April). *What policy makers can learn from school report cards: Analysis of Tennessee's report cards on schools.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco,. CA. (ERIC Document Reproduction Service No. ED 353 265)

Bobbett, G. C., French, R. L., & Achilles, C. M.. (1993, April). *An analysis of report cards on schools: How community/school characteristics impact on student outcome.* Paper presented at the annual meeting of the American Association of School Administrators, Orlando, FL. (ERIC Document Reproduction Service No. ED 368 031)

Bobbett, G. C., French, R. L., & Achilles, C. M.. (1994, November). *Arkansas school district's report cards on schools: What parents, educators, or policy makers can learn from state report cards.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Nashville, TN.

Brown, P. A. (1990). *Accountability in education.* (Contract No. 400-86-0009). San Francisco, CA: Far West Laboratory for Educational Research and Development. (ERIC Document Reproduction Service No. ED 326 949)

Children First Act of 1988. (1988). L3911.C, *Louisiana Revised Statutes.* Baton Rouge, LA: Author.

Connecticut State Department of Education. (1993). *A profile of our schools.* Hartford, CT: Author.

Council for School Performance. (1999). *What is SES and why is it important?* Georgia State University, School of Policy Studies. [On-line]. Available World Wide Web: http://cspweb.gsu.edu/csp/csp_cluster.htm

Council of Chief State School Officers. (1996). *State education accountability reports and indicator reports: Status of reports across the states.* State Education Assessment Center: Washington, DC.:Author.

Franklin, B. J., & Crone, L. J. (1993a, April). *Louisiana progress profiles.* Paper presented at the annual meeting of the American Education Research Association, Atlanta, GA.

Franklin, B. J., & Crone, L. J. (1993b, November). *School accountability: Predictors and indicators of Louisiana school effectiveness.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN.

Kirst, M. W. (1990). *Accountability: Implications for state and local policy makers. Policy perspectives series.* (Report No. IS-90-982). Washington, DC: Office of Educational Research and Improvement.

Kochan, S. E. (1993, November). *How do parents and teachers "grade" Louisiana's school report cards program?* Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 365 727)

Markley, G. (1997, November 14). State schools receive 'C'. *The Opelika-Auburn News.* p. A-1.

Maryland State Department of Education. (1989). *Maryland accountability testing program annual report, 1988-89 school year.* Baltimore: MD.: Author. (ERIC Document Reproduction Service No. ED 314 505)

Mathews, J. G., & Hackett, E. R. (1997, November). *The Alabama superintendent's report card: An analysis of local education agency characteristics and success at meeting state defined performance measures.* Paper presented at the annual meeting of the Mid-South Education Research Association, Memphis, TN.

Mathews, J. G., Ross, M. E., & Spencer, W. A. (1998, November). *An analysis of student achievement using data from the Alabama state report card: Why some schools are successful and others are not.* Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA.

Mississippi State Department of Education. (1993). *Mississippi report card 1993.* Jackson: MS.: Author.

Porter, A. C. (1992). *School delivery standards.* Unpublished manuscript, University of Wisconsin-Madison.

Raivetz, M. J. (1992, April). *Can school districts survive the politics of state testing initiatives?* Paper presented at the annual meeting of the American Education Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 347 680)

Young, B. A., & Smith, T. M. (1997). *The social context of education.* Washington, DC: National Center for Education Statistics. [On-line] Available World Wide Web: http://nces.ed.gov/pubs/yi/index.html

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 1, 51-60

An Empirical Study of Structural Relations Between Science Education and Student Career Aspiration

Jianjun Wang
California State University, Bakersfield

Under pressure of global market competition, science education is a cornerstone supporting development of student career aspiration in high school. A purpose of this study is to disentangle the empirical relationship between career aspiration and science education. Tenth grade data from the National Education Longitudinal Study of 1988 (NELS:88) have been employed to examine factors of educational productivity in three categories, student aptitude-attribute, instructional quantity and quality, and psychological environments pertaining to classroom, home, peer, and news media. Career aspiration is indicated by the nature of student expected jobs, including job stability and academic demand. Direct and indirect links among these factors are cross-examined using the odd- and even-numbered halves of the NELS:88 data. The empirical results reconfirm a strong link between educational attainment and student career aspiration. Through the articulation of educational outcomes, indirect relations are found between career aspiration and contextual factors of educational productivity. Implication of the structural relations has been discussed to facilitate development of career aspiration among the non-college bound students.

The global market competition has made transition between school and workplace more difficult for non-college bound students. In part, the problem is caused by limited school training in core subjects, such as science (Wilson, 1996). Smith et al. (1996) pointed out,

> Competence in science is an important outcome of education. The ability to apply scientific information, interpret data, and make inferences about scientific findings is useful in a world that relies heavily on technological and scientific advances. (p. 74)

While factors of science education attract the attention of many researchers (e.g., Reynolds & Walberg, 1991; 1992), few studies have been conducted to assess the impact of educational productivity on student career aspiration (Wang, 1999). The purpose of this investigation is to examine the empirical relation using 10th grade data from the National Education Longitudinal Study (NELS:88). Thus far, reports of the Third International Mathematics and Science Study (TIMSS) have indicated that the U.S. science score declines in its international standing between grades 8 and 12 (Hunt, 1997). However, tenth grade is a midpoint not included in the TIMSS investigation. In this study, the analysis of NELS:88 data may help disentangle the relationship among factors of science education and student career aspiration.

Significance of the Investigation

Murnane and Levy (1996) stressed that "the most important problem U.S. schools face is preparing children for tomorrow's jobs" (p. 18). While secondary school has a mission of preparing well-trained graduates for higher education, Heckert (1984) cautioned that "The largest segment of the American work force consists of high school graduates who have not attended college, and the nation's economic well-being depends heavily on their performance" (p. xi). As a salient variable in market competition, career aspiration can lead students to pursue better jobs during the school-to-work transition. Raelin (1980) analyzed a national longitudinal database and concluded, "by far the most important attitude uncovered by this study in terms of its contribution to later work experience is career aspiration" (p. 132).

Since the beginning of the 1990s, the end of cold war has enhanced globalization of the labor market. Wilson (1996) noted, "Among the factors that have contributed to the growing gap in employment and wages between low-skilled and college-educated workers is the increased

---

This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation and the National Center for Education Statistics (U.S. Department of Education) under NSF Grant #RED-9452861. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies. Jianjun Wang is Associate Professor of Educational Statistics and Research Design at California State University, Bakersfield. He was formerly research secretary of the China IEA Center and statistical consultant at Kansas State University. Correspondence regarding this article should be addressed to Jianjun Wang, Department of Teacher Education, School of Education, California State University, Bakersfield, 9001 Stockdale Highway, Bakersfield, CA 93311-1099 or by e-mail at jwang@academic.csubak.edu.

internationalization of the U.S. economy" (p. 28). Consequently, some U.S. companies have been lured abroad to seek cheap-labor profit.

Several researchers examined characteristics of disappearing jobs and stressed needs of student attitude preparation (Berliner & Biddle, 1995). Borus (1984) pointed out, "It appears that the specific courses taken in high school, whether academic or vocational, may be less important in determining success in the labor market than other types of learning, such as appropriate work habits and attitudes" (p. 185). The attitude preparation may help retain the cheap labor market and reduce the pressure of unemployment at the non-college bound level. However, the less emphasis on academic training can eventually halt the market upgrading in the United States, and thus, undermine the American welfare. According to Halperin, Melaville, and Taylor (1988), "Regardless of race or ethnicity, the more years spent in education, the greater the annual earnings" (p. 21). The constraint of annual income may inevitably offset student career aspirations for future employment.

An alternative approach is to strengthen academic competency of non-college bound students. Specifically, Bishop (1996) noted, "Increasing numbers of employers need workers who are competent in mathematics, science, technology, and communication" (p. 90). Restricted by the number of school years, enhancement of academic training requires improvement of educational productivity in secondary school.

In the last two decades, researchers have been trying to identify learning factors in school, family, and peer environments (Wang & Staver, 1997). Walberg (1981, 1986) reviewed several thousand theoretical and empirical inquiries in the literature, and developed a nine-factor theory to account for variables of educational productivity. Reynolds and Walberg (1991) elaborated:

> The nine productivity factors can be divided into three sets. First, the student aptitude-attributes set includes (a) student ability or prior achievement, (b) motivation, and (c) developmental level (e.g., age). Second, the instruction set is indexed by its (d) quantity (or amount of time) and (e) quality (or appropriateness) for the student. The third set, psychological environment, includes (f) class environment, (g) the stimulating qualities of the home environment, (h) peer environment, and (i) exposure to mass media, particularly television, outside of school. (p. 97)

Tyler (1984), the forefather of educational assessment, commended, "Herbert Walberg has done a superior interpretation of one of the most massive collections of data on school learning" (p. 27). Despite the existence of a theoretical framework on educational productivity, few researchers examined influence of the productivity factors on career aspiration (Wang, 1999). This study is designed to fill out the literature gap, and disentangle the relationships between career aspiration and factors of science education.

Data Selection

Since the 1970s, the *National Center for Education Statistics* (NCES) has conducted three longitudinal surveys at the non-college bound level, *The National Longitudinal Study of the 1970s* (NLS-72), *High School and Beyond of the 1980s* (HS&B), and *The National Education Longitudinal Study of 1988 (NELS:88)*. NELS:88 is the most recent project building on the experiences from the NLS-72 and HS&B studies (Davis & Sonnenberg, 1995). One of the goals of NELS:88 is to examine the transition of secondary students to the work force (Ingels et al., 1994). Accordingly, career aspiration is an indispensable component in the NELS:88 investigation.

The NELS:88 data cover three school levels, 8th, 10th and 12th grades. The 10th grade data have been chosen in this study for the following reasons. First, students in the 8th and 12th grades are subjected to adjustment for entry to and graduation from high school, which may interfere with student career aspiration. In contrast, 10th graders are in the middle of a relatively stable period in which they can focus more on academic learning. Second, "dropouts from the 1990 sophomore class were more likely to return to school than were their counterparts a decade earlier" (Smith et al., 1996, p. 50). Thus, the sophomore level is a crucial stage at which student career aspiration may play a role in the transition from school to work place.

The 10th grade data were collected in 1990 during the first follow-up survey of NELS:88. In the 1990s, the end of the cold war facilitated further development of the global labor market. Wilson (1996) asserted, "In the new global economy, highly educated and thoroughly trained men and women are in demand" (p. 29). With rapid development of information technology, the global job market continues the momentum of expansion in the 21st century. In this context of ongoing market globalization, the 10th grade data collected in the early 1990s are still relevant to an examination of student career aspiration.

In educational psychology, career aspiration is a complex construct with two interrelated components, *job acquisition* and *career stability* (Payne & Peck, 1979; Spiessl, 1981). The NELS:88 project addressed both components in its 10th grade survey. The component of job acquisition is represented by educational demand on student-expected future jobs (Item ID: F1S66D) and career stability is reflected by student-perceived importance of finding a steady job (Item ID: F1S46E). The two items were incorporated in the NELS:88 survey questionnaire and coded on Likert type scales. The questions were stated as:

> Do you agree with the following statement about why you go to school?
>
> . *Education is important for getting a job later on* (Item ID: F1S66D)
>
> strongly agree agree disagree strongly disagree
>
> How important is the following item to you in your life?
>
> . *Being able to find steady work* (Item ID: F1S46E)
>
> not important somewhat important very important

The indicators of job stability and educational demand are generally applicable to either gender group because "both men and women with less education were less likely to work consistently than those with higher levels of education" (Clery, Lee, Knapp, & Carroll, 1998, p. 45). In addition, great effort has been made by the NELS:88 researchers to strengthen reliability of the survey outcomes on the ethnicity dimension. Rock, Pollack, and Quinn (1995) reported, "the NELS:88 battery was specifically designed to reduce the gap in reliabilities that is typically found between the majority group and the racial/ethnic minority groups" (p. 4).

In summary, the NELS:88 data are pertinent to a study of career aspiration. In addition, Johnson (2000) noted that the NELS:88 data can be employed to analyze factors of educational productivity. On the basis of the rich information collected in NELS:88, statistical analyses were conducted to examine the relationship between student career aspiration and factors of educational productivity.

Research Questions

Due to the limited schooling at the non-college bound level, raising educational productivity is an important approach to enhancement of student career aspiration. Halperin, Melaville, and Taylor (1988) observed, "Youth today, especially those who do not go to college, find it increasingly difficult to match changing market demands" (p. 7-8). Accordingly, research questions investigated in this study are:

1. What are the empirical relations among factors of educational productivity and student career aspiration?
2. Is the model of empirical relations strongly supported by the NELS:88 data base?
3. What interpretations can be made on the structural relations in terms of the existing theory of educational productivity?

Methods

Factors of Educational Productivity

Since the indicators of career aspiration reflect student determination on pursuing education-based, stable jobs in the future, development of career aspiration is closely linked to factors of educational productivity. Walberg's (1981) productivity theory presents a useful framework guiding identification of the empirical factors in science education. Young, Reynolds, and Walberg (1996) recollected:

> The theory has guided the compilation of more than 120 research syntheses of 8,000 comparisons in small-scale experimental and correlational studies (Fraser, Walberg, Welch, & Hattie, 1987) and 23 regression analyses of achievement obtained from (mostly national) surveys of about 250,000 students in six subjects of primary and secondary school study. (Paschal & Starhia, 1992, p. 272)

In the original Walberg theory, student age has been included as a factor of educational productivity. In analyzing a 10th grade data set, Reynolds and Walberg (1992) noted that "Because the students in the sample were all from the same grade level, age was relatively constant and therefore omitted" (p. 373). A similar situation exists in the NELS:88 data analysis. Due to the selection of students from the same grade, approximately 93% of the 10th graders were born in 1973 and 1974. Therefore, the NELS:88 data have little variation on the age dimension.

Guided by Walberg's (1981) theory, the first group of productivity factors is classified in an aptitude-attribute category, including "(a) student ability or prior achievement, (b) motivation, and (c) developmental level (e.g., age)" (Reynolds & Walberg, 1991, p. 97). Student ability or prior achievement is part of the educational outcome resulting from individual commitment (Walberg, 1981). In this study, *educational outcome* is represented by student overall science proficiency, as well as student-perceived importance of school grades. In addition, *motivation* is indicated in NELS:88 by student-reported reluctance on cutting classes and being late for school. The age factor is omitted due to its slight variation in the 10th grade data. A Cronbach's α coefficient was computed to assess reliability of the aptitude-attribute scale, and the result is $\alpha = 0.50$.

The second group of instructional variables "is indexed by its (d) quantity (or amount of time) and (e) quality (or appropriateness) for the student" (Reynolds & Walberg, 1991, p. 97). Specifically, *quantity* is assessed by the homework time students spent in and out of school. *Quality* is represented by student responses on the effectiveness of teaching and teacher-student communication. Cronbach's α coefficients for the *quantity* and *quality* scales are 0.50 and 0.66, respectively.

The third group variables address student learning environment. According to Walberg's theory, the learning environment has four aspects, *home*, *class*, *peer*, and *media influence* (Reynolds & Walberg, 1991, 1992). Home environment is described by socio-economic status and parental education (Cronbach's α=0.93). Class environment is indicated by students' responses on whether they feel put down by teachers and classmates (Cronbach's $\alpha = 0.50$). Friends who value study and good grades are indicators of a positive peer environment (Cronbach's $\alpha = 0.76$). Finally, mass media influence was represented by the number of hours students spent on TV watching over weekdays and weekends (Cronbach's $\alpha = 0.71$).

Given the multiple aspects of educational productivity, not all the factors are easily identifiable through survey questionnaires. To facilitate an empirical study of various factors in science education, the measurement scales with moderate to high reliability indexes have been employed in this study to articulate relations between career aspiration and educational productivity (Table 1). Results of the quantitative inquiry will be further discussed qualitatively to disentangle meaning of the statistical findings in a school setting.

Model Confirmation

To cross-examine results of the statistical analyses, the NELS:88 data are split into even- and odd-numbered halves by selecting every other case. Cases in the odd-numbered half were employed to develop a structural equation model, and the even-numbered half was adopted to cross-validate the findings between career preparation and educational productivity.

In 1990, a total of 16,581 tenth graders was chosen randomly from the U.S. to take a science test. To construct correlation matrices among the selected factors, pairwise deletions were implemented at the stage of data cleaning. As a result, the minimum size of the retained sample is 14,720. The retention rate of 89% is above the standard of 70% rate set by NCES (Kaufman et al, 1999).

Table 1
Description of Latent Factors and the Corresponding Nels:88 Indicators

Factor	NELS:88 Indicator	Description
Career Orientation	F1S66D	Education is important to get a job later
	F1S46E	Importance of being able to find steady work
Education outcome	F12XSPRO	Overall science proficiency
	F1S38	Importance of being able to get good grades
Motivation	F1S12A	Feel ok to be late for school
	F1S12B	Feel ok to cut a couple of classes
Instructional quantity	F1S36C1	Time on science homework in school
	F1S36C2	Time on science homework out of school
Instructional quality	F1S7G	The teaching is good at school
	F1S7I	Teacher praise efforts of working hard
	F1S7L	Most teachers listen to me
Home environment	F1SES	Home socioeconomic status
	F1PARED	Parents' highest education level
Class environment	F1S7J	Feel put down by teachers in class
	F1S7K	Feel put down by students in class
Peer environment	F1S70B	Among friends, study is important
	F1S70D	Among friends, good grades are important
Mass media	F1S45A	Weekday watching TV hours
	F1S45B	Weekend watching TV hours

Bentler and Bonett (1980) cautioned that the probability of detecting a false model increased with the sample size. To guard against the potential type I error, the minimum sample size is used in the LISREL program for parameter estimation. The structural equation model is developed on the basis of the literature on career aspiration and Walberg's theory of educational productivity. Direct structural relations are postulated between career aspiration and factors of educational productivity, because Walberg's model posits direct, simultaneous influences of the nine factors of educational productivity on schooling outcomes (Reynolds & Walberg, 1991; 1992).

Results

Path coefficients between factors of educational productivity and career aspiration are presented in Figure 1. Because the NELS:88 data were split into halves, results on the even numbered half are enclosed within boxes. Data in the odd-numbered half converge to the results after 100 iterations, and the even-numbered half complete the convergence in 241 iterations.

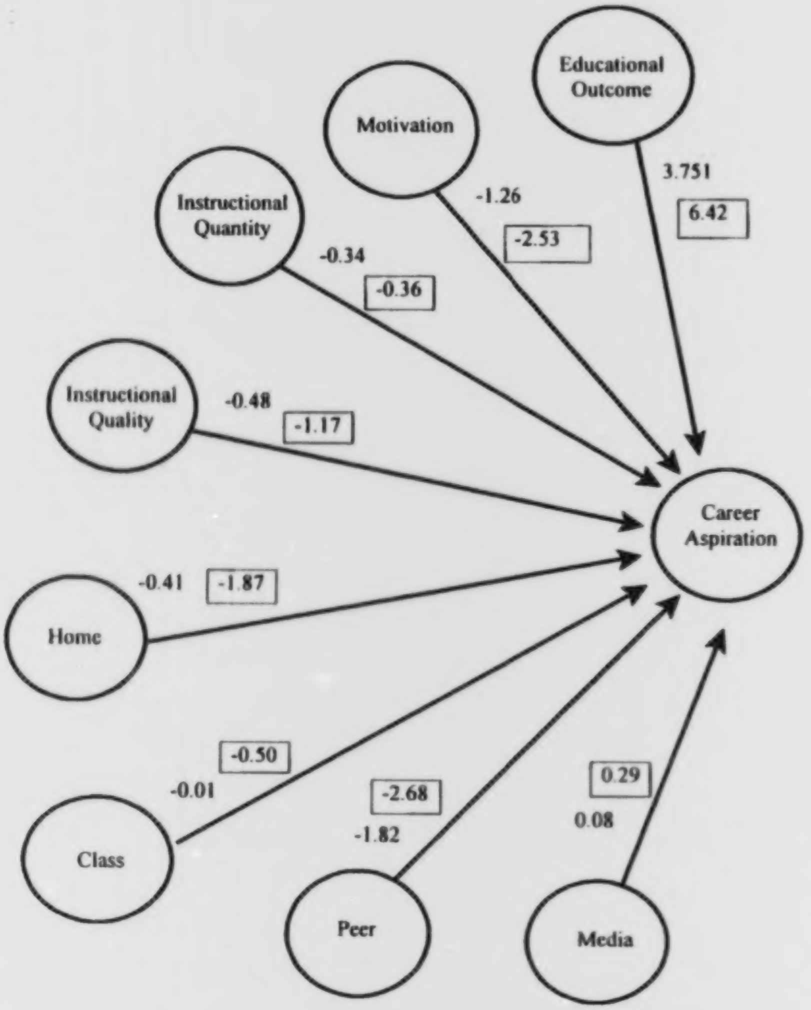


Figure 1. Parameters Describing the Direct Effects of Educational Productivity on Career Aspirations

Not shown in Figure 1 to simplify the presentation are inter-relationships among the eight factors of educational productivity. In fact, correlations among the eight factors of educational productivity have been considered in this investigation (Table 2). The maximum likelihood method is employed to estimate *direct* structural relations in Figure 1. Potential *indirect* relations were examined in the path analysis using $p_{indirect} = rp_{direct}$ (Loehlin, 1992), where r is the correlation coefficient between the direct and indirect factors (Table 3). To balance the effect of variable scaling, all indicators are standardized with large values representing positive responses.

Selection of the model-fitness indicators was also grounded on the research literature. Marsh, Balla, and McDonald (1988) suggest the use of root mean square residual (RMR) to "justify the conclusion that a model adequately fits a particular set of data" (p. 391). LIREL's goodness-of-fit index (GFI) and adjusted goodness-of-fit index (AGFI) are also recommended for use to assess potential gaps between the proposed model and the empirical data (Joreskog, & Sorbom, 1993). In addition, the incremental fit index (IFI) is computed to examine the goodness of fit independent of sample sizes. Bentler and Bonett (1980) observed,

> Authors of literature on psychological statistics are well aware of the distinction between statistical significance and practical significance; an incremental fit index can provide information about practical significance, in which a statistically significant effect can be evaluated for its practical usefulness in explaining the data. (p. 599)

Table 2
Correlations of the productivity factors in the structural equation model

	ed outcome	motivation	quantity	quality	home	class	peer	media
Odd-ID Half								
ed outcome	1.00							
motivation	0.72	1.00						
quantity	0.28	0.11	1.00					
quality	0.56	0.40	0.12	1.00				
home	0.16	-0.05	0.11	0.03	1.00			
class	0.34	0.24	0.06	0.57	0.01	1.00		
peer	0.81	0.44	0.16	0.35	0.04	0.18	1.00	
media	-0.10	-0.04	-0.08	-0.06	-0.16	-0.03	-0.04	1.00
Even-ID Half								
ability	1.00							
motivation	0.70	1.00						
quantity	0.13	0.03	1.00					
quality	0.60	0.43	0.05	1.00				
home	0.34	-0.04	0.04	0.04	1.00			
class	0.33	0.17	0.02	0.45	0.02	1.00		
peer	0.74	0.42	0.07	0.36	0.07	0.14	1.00	
media	-0.15	-0.04	-0.02	-0.04	-0.14	-0.01	-0.07	1.00

Table 3
Indirect path coefficients from productivity factors to career aspiration through education outcome

	P_{direct}	Indirect path through education outcome to career aspiration ($p_{indirect}=rp_{direct}$)						
Sample	ed outcome	motivation	quantity	quality	home	class	peer	media
Odd-ID Half	3.75	2.70	1.05	2.10	0.60	1.28	3.04	-0.38
Even-ID Half	6.42	4.49	0.83	3.85	2.18	2.11	4.75	-0.96

All these model-fit indexes have been presented in Table 4. The use of multiple goodness-of-fit indicators is strongly recommended by Bollen (1989).

Table 4
Goodness-of-fit Indicators for the Structural Equation Model

Sample	RMR	GFI	AGFI	IFI
Odd-ID Half	0.044	0.97	0.96	0.95
Even-ID Half	0.039	0.98	0.96	0.95

For each productivity factor, factor loadings have been calculated to reflect the contribution of the NELS:88 indicators on the latent factor construction (Table 5).

Table 5
Description of Latent Factors and the Corresponding Loadings of the Nels:88 Variables

		Factor Loading	
Factor	NELS:88 Variable	Odd-ID Half	Even-ID Half
Career orientation	F1S66D	0.59	0.59
	F1S46E	0.31	0.31
Education outcome	F12XSPRO	0.12	0.20
	F1S38	0.60	0.51
Motivation	F1S12A	0.63	0.63
	F1S12B	0.76	0.75
Instructional quantity	F1S36C1	0.28	0.09
	F1S36C2	1.24	3.23
Instructional quality	F1S7G	0.60	0.57
	F1S7I	0.61	0.58
	F1S7L	0.71	0.67
Home environment	F1SES	0.98	0.93
	F1PARED	0.86	0.91
Class environment	F1S7J	0.97	1.08
	F1S7K	0.35	0.29
Peer environment	F1S70B	0.77	0.78
	F1S70D	0.80	0.79
Mass media	F1S45A	1.34	1.50
	F1S45B	0.43	0.38

Discussion

One of the goals for improving educational productivity is to facilitate student school-to-work transition (Murnane & Levy, 1996). At the non-college bound level, most students have difficulty finding a steady job right after graduation (Bishop, 1996). In addition, educational demand of the expected jobs increases with development of modern technology (Decker, 1997). To help nurture student career aspiration, results of the quantitative inquiry are discussed in this section to add relevant qualitative perspectives to the statistical findings.

As was shown in Table 1, job stability (Item ID: F1S46E) and education demand (Item ID: F1S66D) are indicators of career aspiration. The NELS:88 data analysis reveals a higher factor loading on the aspect of educational demand (see Table 5). To a certain extent, the more weight on educational demand has reflected the importance of adequate academic training for future employment. Wilson (1996) recollected, "The unemployment rates among both low-skilled men and women are five times that among their college-educated counterparts" (p. 28). Thus, the job market competition demands solid academic training, and improvement of educational productivity can be an effective approach to supporting development of career aspirations among the non-college bound students.

By the same token, Figure 1 shows a strong link between career aspiration and educational outcome. Between two indicators of educational outcome (see Table 1), the *overall science proficiency* (Item ID: F12XSPRO) carries less weight than the *importance of getting good school grades* (Item ID: F1S38) (see Table 5). In part, this finding is resulted from the indicator selection. Specifically, as an indicator of educational outcome, the importance of school grade confirms the educational demand for career aspiration (F1S66D). Decker (1997) reviewed comparative statistics in education and found that "The industrialized countries with the highest productivity levels tend to have highly educated work forces, and the convergence in productivity among these countries generally parallels that in educational attainment" (p. 5). This parallelism is corroborated by the confirmatory result of the NELS:88 data analysis.

In contrast, other aspects of educational productivity are focused on background factors including *student motivation, instructional quantity and quality, home resources, classroom environment, peer grouping* and *media influence* (Table 1). Indicators of these factors are systematically recoded with a large value representing a positive response. The positive factor loadings in Table 5 indicate that the positive scaling is in line with construction of the latent factors.

In addition, differences in the values of factor loading (Table 5) seem to suggest that not all the indicators carry the same weight on the latent factor construction. For instance, the factor of instructional quantity is indicated by the amount of time students spent on science homework in and out of school. The factor loading is much higher for the out-of-school time. While the in-school time can be controlled by instructors, the out-of-school time can vary among students, depending on their personal and parental commitment to education. Consequently, the amount of homework time at home is important to completing the school work.

Inspection of Table 2 also suggests a weak correlation ($r \leq 0.11$) between instructional quantity and home environment. In fact, the factor of home environment has weak correlations with almost all productivity factors except for educational outcome (see Table 2). Perhaps the results depict that most parents care more about learning outcomes. Therefore, much less attention has been paid to the instructional factors, student motivation, as well as the class, peer, and media environments.

The media factor is indicated by the number of hours students spent watching TV over weekdays and weekends. The weekday hours have a much higher factor loading than the weekend hours. Partly because the TV watching may have taken time from course work, the media factor has negative correlations with all other factors of educational productivity (see Table 2). It appears that proper control of TV watching, particularly on weekdays, may help raise the level of educational productivity in school.

The school setting is described by class environment. Indicators of this factor reflect student feeling of being put down by teachers and classmates (Table 1). The impact from teachers is much more important in terms of the factor loading (Table 5). Meanwhile, the class factor shows a very weak correlation with the quantity of instruction ($r \leq 0.06$). It seems that further instructional improvement can be made if teachers are more sensitive to student feelings and take proper steps to articulate the teaching effort with pedagogical skills.

Besides the aforementioned weak correlations, the remaining factors are linked extensively through correlations (Table 2). Thus, given the few exceptions, results of the NELS:88 data analysis generally concurs with the empirical findings of Fraser, Walberg, Welch, and Hattie (1987), i.e., "all factors seem important in that, without at least a small amount of each, students are likely to learn little" (p. 227).

In addition to the direct links between factors of educational productivity and career aspiration (Figure 1), indirect effects may emerge through the factor of *educational outcome* to connect other productivity factors with student career aspiration. The NELS:88 results indicate that the indirect path coefficients are consistently larger than the corresponding direct path coefficients (see Table 3 vs. Figure 1). Hence, *educational outcome* may have served as an important channel passing the indirect influence of productivity factors onto student career aspiration.

Despite the salient role of educational outcome, Ball and Goldman (1997) observed,

> Schools almost never ask corporations what they want future employees to know, nor do they survey their graduates about how their schooling affected their careers. . . . For their part, businesses rarely help schools develop their courses or define necessary skills. (p. 231)

Therefore, to smooth the school-to-work transition, further collaboration between educators and business leaders may help enhance student career aspirations in the global market competition.

While the indirect links are positive for most productivity factors (Table 3), media influence appears to be the only exception. On one hand, Dynan and Fraser (1985) noted, "Outside of school, students are exposed to many media from which to learn science concepts" (p. 1). The direct path coefficient in Figure 1 seems to confirm the positive link between media influence and career aspiration. On the other hand, educators are concerned about the media influence that may have negatively distracted student attention from academic course work. Bishop (1996), for instance, suggested that "Parents must tell children, 'Turn off the TV and do your homework'" (p. 80). Since *media influence* was specifically represented in this study by the number of hours students spent watching TV (Table 1), the distractive feature seems to have predominated the indirect path coefficients via *educational outcome* (Table 3).

Table 4 presents indexes of goodness-of-fit between the NELS:88 data and the empirical model linking educational productivity and career aspiration. The small values of the root mean square residual (RMR $\leq$ 0.044) indicate an adequate fit of the model with the NELS:88 data. This conclusion is also supported by the high goodness-of-fit index (GFI $\geq$ 0.97) and adjusted goodness-of-fit index (AGFI = 0.96). Joreskog and Sorbom (1981) pointed out that GFI "is a measure of the relative amount of variances and covariances jointly accounted for by the model" and AGFI "corresponds to using mean squares instead of total sums of squares" (p. I.40-41). Hence, AGFI is based on a penalty function for additional parameters. Given the fact that so many factors are involved in this study, the small gap between GFI and AGFI seems to suggest little room for model improvement through different parameterizations.

It should be noted that the data cleaning process has resulted in different sample sizes for the selected variables, and the minimum sample size is employed to avoid detecting a false model (Bentler & Bonett, 1980). Marsh, Balla, and McDonald (1988) noted that incremental fit indexes "are useful for comparing the fit of a particular model across samples that have unequal sizes" (p. 393). The incremental fit indexes are 0.95, concurring the model-fit conclusion suggested by RMR, GFI, and AGFI (see Table 4).

In summary, the NELS:88 data analysis indicates that not all productivity factors carry equal weight on enhancing student career aspiration. The entire system of education appears rather complex, accommodating *direct* and *indirect* links among many contextual factors. On the basis of the factors suggested by Walberg's productivity model, *educational outcome* plays a pivotal role of articulating educational productivity and career aspiration. Multiple indicators of goodness-of-fit consistently suggest an adequate fit of the empirical model to the NELS:88 data. The empirical results have been discussed to facilitate further enhancement of the link between educational productivity and student career aspiration at the non-college bound level.

References

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Berliner, D. & Biddle, B. (1995). *The manufactured crisis*. New York, NY: Addison-Wesley.

Bishop, J. (1996). Singling the competencies of high school students to employers. In L. Resnick and J. Wirt (Eds.), *Linking School and Work*. San Francisco, CA: Jossey-Bass.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Borus, M. (1984). *Youth and the labor market*. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Clery, S. B., Lee, J. B., Knapp, L. G., & Carroll, C. D. (1998). *Gender differences in earnings among young adults entering the labor market*. Washington, DC: U.S. Department of Education.

Davis, C. & Sonnenberg, B. (1995). *Programs and plans of the National Center for Education Statistics*. Washington, DC: U.S. Department of Education.

Decker, P. (1997). *Findings from education and the economy: An indicators report*. Washington, DC: U.S. Department of Education.

Dynan, M. B., & Fraser, B. J. (1985). *Informal learning of science*. Perth, Australia: Curtin University of Technology. (ERIC Document Reproduction Service No. ED 340 559).

Fraser, B., Walberg, H., Welch, W., & Hattie, J. (1987). Conclusions and implications. *International Journal of Educational Research, 11*, 227-235.

Halperin, S., Melaville, A., & Taylor, A. (1988). *The forgotten half: Non-college youth in America*. Washington, DC: The William T Grant Foundation Commission on Work, Family and Citizenship.

Heckert (1984). Why another education report? In Committee on Science, Engineering, and Public Policy (Eds.), *High schools and the changing workplace*. Washington, DC: National Academy.

Hunt, J. B. (1997). *The national education goals report: Summary*. Washington, DC: National Education Goals Panel.

Ingels, S. J., Dowd, K. L., Baldridge, J. D., Stipe, J. L., Bartot, V. H., Frankel, M. R., & Quinn, P. (1994). *Second follow-up: Student component data file user's manual* (NCES 94-374). Washington, DC: U.S. Department of Education.

Johnson, R. (2000, April). *Walber's educational productivity model and the NELS:88 database*. Paper presented at the annual meeting of American Educational Research Association, New Orleans, LA.

Joreskog, K. G., & Sorbom, D. (1981). *Analysis of linear structural relations by the method of maximum likelihood*. Chicago, IL: International Educational Services.

Joreskog, K. & Sorbom, D. (1993). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.

Kaufman, P., Chen, X., Choy, S. P., Ruddy, S. A., Miller, A. K., Chandler, K.A., Chapman, C. D., Rand, M. R., & Klaus, P. (1999). *Indicators of school crime and safety, 1999*. (NCES 1999-057/NCJ-178906). Washington, DC: U.S. Departments of Education and Justice.

Loehlin, J. C. (1992). *Latent variable models* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410.

Murnane, R. & Levy, F. (1996). *Teaching the new basics: principles for educating children to thrive in a changing economy*. New York, NY: Free Press.

National Education Longitudinal Study of 1998 (NEL:88). [CD-ROM]. (1996). Available: National Center for Education Statistics, Washington, D.C. [1996, May 28].

Payne, G. N. & Peck, R. F. (1979). *Coping strategies, career aspiration, and career values in the career decision-making process*. [No city information], TX: [No institution information]. (ERIC Document Reproduction Service No. ED189 312).

Raelin, J. (1980). *Building a career*. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Reynolds, A. J. & Walberg, H. J. (1991). A structural model of science achievement. *Journal of Educational Psychology, 83*(1), 97-107.

Reynolds, A. J. & Walberg, H. J. (1992). A structural model of science achievement and attitude: An extension to high school. *Journal of Educational Psychology, 84*(3), 371-382.

Rock, D. A., Pollack, J. M., & Quinn, P. (1995). *Psychometric report for the NELS:88 base year through second follow-up (NCES 95-382).* Washington, DC: U.S. Department of Education.

Smith, W. S. (1982). Career education attitudes and practices of K-12 science educators. *Journal of Research in Science Teaching, 19*(5), 367-375.

Smith, T. M., Young, B. A., Choy, S. P., Perie, M., Alsalam, N., Rollefson, M. R., & Bae, Y. (1996). *The condition of education 1996* (NCES 96-304). Washington, DC: U.S. Department of Education.

Spiessl, R. W. (1981, October). *An adaptation of dual labor market theory to the evaluation of a youth employment project.* Paper presented at Joint Annual Meeting of the Evaluation Network and Evaluation Research Society, Austin, TX.

Tyler, R. W. (1984). A guide to educational troubleshooting. *Educational Leadership, 41*(8), 27-30.

Walberg, H. (1981). A psychological theory of educational productivity. In F. H. Farley and N. Gorden (Eds.), *Psychology and education.* Berkerley, CA: McCutchan.

Walberg, H. (1986). Synthesis of research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). Washington, DC: American Educational Research Association.

Wang, J. & Staver, J. (1997). An empirical study about Chinese adolescent science achievement: gender differences. *Journal of Educational Research, 90*(4), 252-255.

Wang, J. (1999). A structural model of student career aspiration and science education. *Research in Schools, 6*(1), 53-63.

Wilson, J. W. (1996). *When work disappears.* New York, NY: Alfred A. Knopf.

Young, D., Reynolds, A., & Walberg, H. (1996). Science achievement and educational productivity: A hierarchical linear model. *Journal of Educational Research, 89*(5), 272-278.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 1, 61-75

The Role of Cooperative Learning in Research Methodology Courses: A Mixed-Methods Analysis

Anthony J. Onwuegbuzie
Valdosta State University

Denise A. DaRos-Voseles
Youngstown State University

This study investigated the effectiveness of cooperative learning (CL) in a graduate-level research methodology course. Participants comprised 193 graduate students enrolled in several sections of this course. Eighty-one students were in sections wherein CL groups were formed to undertake the major course requirements; 112 were in sections wherein all assignments were undertaken individually (IL). Students' conceptual knowledge of research concepts, methodologies, and applications was measured individually in both groups via midterm and final examinations. A split-plot analysis of variance revealed a group by examination time interaction, whereby CL students had statistically significantly lower performance levels on the midterm examination than did IL students (effect size = 0.48). However, no statistically significant difference in achievement was found with respect to the final examination. Analysis of reflexive journals indicated that most students (70.4%) tended to have positive overall attitudes towards their cooperative learning experiences. Implications are discussed.

Overview of Cooperative Learning

To date, cooperative learning is one of the most thoroughly researched of all instructional methods (Slavin, 1992). Over the last several years more than 1,000 studies have been conducted at the precollegiate level (Cooper & Muech, 1992). The work of David and Roger Johnson at the University of Minnesota, Robert Slavin at Johns Hopkins University, and Elizabeth Cohen at Stanford University emerges most frequently. They are considered the pioneers in cooperative learning, having devoted years of detailed research and analysis to clarify the conditions under which cooperative, competitive, and individualized goal structures influence student achievement, self-esteem, and social skills (Smith & MacGregor, 1992).

Numerous benefits emerged from the flurry of cooperative learning research and the associated publication of three major works (Johnson & Johnson, 1989; Sharan, 1990; Slavin, 1990). These publications expanded the findings of earlier reviews and provided further evidence that cooperation improved self-esteem, increased effort to achieve, enhanced psychological well-being and caring relationships, and fostered the ability to take the perspective of another individual (Davidson & Kroll, 1991).

Cooperative learning is defined as "the instructional use of small groups so that students work together to maximize their own and each other's learning" (Johnson, Johnson, & Smith, 1991a, p. III). This instructional strategy is based on the social interdependence theories of Morton Deutsch and Kurt Lewin (Deutsch, 1949; Lewin, 1935). These theories and related research suggest that social interdependence positively influences individual interaction with a given situation, which subsequently affects the outcomes of that interaction (Johnson & Johnson, 1989).

Johnson et al. (1991a) assert that grouping students to work on a common assignment does not guarantee that co-operative learning will occur. Apparently, this learning method requires more structured small group learning around specifically defined tasks or problems. And although there are a variety of ways to implement cooperative learning in any discipline, researchers (Johnson, Johnson, & Holubec, 1991; Johnson et al., 1991a; Johnson, Johnson, & Smith, 1991b) have identified five essential elements: positive interdependence, face-to-face promotive interaction, individual accountability, social skills, and group processing.

The first requirement for a successful cooperative activity is positive interdependence among group members. That is, each group member's efforts are indispensable and necessary for the groups' success, and each member has a distinctive contribution to make to the

---

Anthony J. Onwuebuzie is an Assistant Professor in the Educational Leadership Department at Valdosta State University (Georgia). Denise DaRos-Voseles is an Associate Professor in the Department of Teacher Education at Youngstown State University. Correspondence regarding this article should be addressed to: Anthony Onwuegbuzie, Department of Educational Leadership, College of Education, Valdosta State University, Valdosta, Georgia, 31698 or by e-mail to tonwuegb@valdosta.edu.

group effort. In tandem with positive interdependence is face-to-face promotive interaction among group members. The structuring of tasks enables each others' strengths and weaknesses to be used in a complimentary manner to reach the group's goals. Efficiently and effectively helping one another and exchanging needed resources are important strategies used to reach these goals (Johnson, Johnson, & Holubec, 1991).

Individual accountability places the responsibility on the student to master the assigned work. In so doing, *coattailing* (i.e., disproportionately benefitting from another's work) is less likely to occur. Cooperative learning groups provide a forum to use social skills such as effective communication, building and maintaining trust, and constructively resolving conflicts. In fact, Johnson and his colleagues (Johnson, 1993; Johnson & Johnson, 1991; Johnson et al., 1991a, 1991b) contend that the success of a group's work is contingent on healthy interaction between students. Lastly, groups must assess how well they are working towards achieving their goals. Johnson and Johnson (1991) postulate that the aforementioned five elements help to insure a successful cooperative learning experience for students. Notably, these elements are an amalgam of interpersonal skills and learning outcomes.

Smith, Johnson, and Johnson (1992) theorized that there are a variety of cooperative learning activities which can be classified into the following three group types: informal learning groups, formal cooperative learning groups, and cooperative base groups. According to this conceptualization, informal learning groups are less structured and short-term, requiring students to complete a task often associated with a lecture. Formal cooperative learning groups are longer in duration, comprise small (2-4 member) groups, and are established by the instructor to create a final product such as a course project. Cooperative base groups are stable, long-term, peer support groups composed of 3-5 individuals. According to Smith et al. (1992), base groups enhance students' learning and increase attendance in larger lecture classes. Learning becomes a melding of task-oriented processes and interpersonal experiences that enhance valuable teaming skills.

Summary of Cooperative Learning Efficacy

The efficacy of cooperative learning has been studied at all age levels for many subject areas, in all types of school settings, and with students of different ethnic groups (Johnson et al., 1991a; Slavin, 1991a). Slavin (1990) identified more than 70 high-quality studies that compared learning and traditional methods in elementary and secondary schools. Of the 70 studies, 67 measured effects on student achievement, with 41 (61%) reporting significantly higher achievement levels in cooperative than in control classes. Twenty-five (37 percent) found no differences, and in only one study did the control group outperform the experimental group.

Meta-analytic Studies of Cooperative Learning Research

Johnson and colleagues (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981) conducted a meta-analysis of 122 achievement studies. The majority of these studies measured several outcomes in addition to achievement. Their findings supported the effectiveness of cooperative learning in a variety of forms. Not only did achievement levels increase; so did levels of self-esteem, attitudes towards school, time-on-task, and attendance rate. It is thus not surprising that cooperative learning as an instructional method was recommended by the National Council of Teachers of Mathematics (1989, 1991) and the National Research Council (1989).

Johnson and Johnson (1989) later conducted a more extensive meta-analysis of over 575 experimental and 100 correlational studies (see Johnson & Johnson, 1989 for a complete listing of these studies). These studies spanned 90 years with different age subjects, in different subject areas, and in a variety of settings. A subgroup of 375 studies was reviewed to answer the question of how successful competitive, individualistic, and cooperative efforts were in promoting productivity and achievement. The achievement effects found were as follows:

> When all of the studies were included in the analysis, the average student cooperating performed at about two-thirds a standard deviation above the average student learning within a competitive (effect size= 0.67) or individualistic (effect size = 0.64) situation. When only high-quality studies were included in the analysis the effect sizes were 0.88 and 0.61 respectively (Johnson et al., 1991a, p. 38).

Although many studies in the area of cooperative learning have been conducted with students in Grades 3-9 (Purdom & Kromey, 1992), relatively few studies have examined the effects of this method of instruction in Grades 10-12 (Newman & Thompson, 1987). In fact, reflecting this trend, the only full-scale review that focused solely on secondary schools (i.e., middle, junior, and high schools) was conducted by Newman and Thompson (cited in Slavin, 1992). Their review on cooperative learning located 27 reports of what they deemed to be high-quality studies, including 37 comparisons of cooperative versus control methods. However, consistent with studies in earlier grades, a statistically significant proportion of these studies (68%) supported the use of cooperative learning methods.

Overview of Cooperative Learning in Higher Education

Even fewer studies have been conducted at the collegiate level (Slavin, 1989, 1991b). Unfortunately, of those that have, many suffer from a variety of methodological problems (Cooper & Muech, 1992). Nevertheless, Slavin (1992) posited that, despite flawed research at this level, there is evidence of positive achievement effects of this learning strategy in college settings (see, for example, Fraser, Beaman, Diener, & Kelem, 1977). A subgroup of the earlier noted 1989 meta-analysis was conducted by Johnson and Johnson (1989). This subgroup of 137 experimental studies compared cooperative, competitive, and individualistic efforts at the college and adult levels, revealing that cooperative learning promotes higher achievement than do competitive or individualistic learning (effect sizes = 0.59 and 0.62, respectively) (Johnson et al., 1991a). Qin, Johnson, and Johnson (1995), in a review of 46 studies at the post-secondary level, found positive effects on problem solving associated with the cooperative learning model in 55 of the 63 outcomes. Given this caveat, there are still many questions to be answered; notably why and how cooperative learning yields its various effects at the college level. To this end, Slavin (1989) recommended that research continue to add to the knowledge base in higher education.

Although the research on cooperative learning indicates that it is a powerful method for influencing student achievement and attitudes, this technique has not been implemented systematically in many college classrooms (Cooper & Muech, 1992). Advocating for this instructional method, Garfield (1993) provides an in-depth rationale for the use of cooperative learning in college statistics classes. Indeed, results of studies are beginning to emerge on the use of cooperative learning activities in statistics courses (Borresen, 1990; Giraud 1997; Jones, 1991; Keeler & Steinhorst, 1995; Shaughnessy, 1977). In particular, instructors of undergraduate introductory and applied statistics classes have examined the effects on achievement of lecture versus cooperative teaching methods (Borresen, 1990; Giraud, 1997; Keeler & Steinhorst, 1995). Teacher-researchers in these studies acknowledged that many students have negative attitudes towards statistics; therefore, in addition to performance outcomes, the question of whether or not cooperative learning techniques can change negative attitudes was addressed. The length of these studies ranged from two sections in the same semester (Giraud, 1997) to three years comprising six introductory classes (Borresen, 1990).

Borresen (1990) postulated that students in voluntary cooperative learning groups would outperform students in assigned cooperative learning groups, and that both groups would perform better than students working individually. Findings revealed that more than twice as many "A's" were earned by the assigned and volunteer groups than by students working on their assignments individually. Interestingly, no achievement difference was found between the voluntary and assigned groups. Of equal concern to the researcher was students' attitudes towards statistics. Based on formal and informal evaluations, approximately half of the students in the voluntary and assigned groups indicated a positive change in their attitudes and enthusiasm.

Giraud (1997) examined the comparative effects of cooperative versus lecture methods of instruction in two sections of an undergraduate statistics course. Giraud randomly assigned students of various abilities to create opportunities for supportive scaffolding (Vygotsky, 1978). The scaffolding process occurs when less skillful students actively collaborate with more competent peers, thereby enabling students to develop more complex levels of understanding and skill. The size of the cooperative groups ranged from three to five students and remained relatively consistent throughout the semester. A 30-item Statistics Readiness Test was administered during the first week of class to test students' knowledge of basic algebra and mathematical reasoning. According to Giraud, this instrument measured statistical readiness and was used to assist in determining whether there was an interaction between students' basic algebra and mathematics skill level and the instructional method. The major question posed was whether or not cooperative learning benefitted less skilled students. Post hoc comparisons revealed that students who scored the lowest on the algebra and mathematics reasoning pretest and were in the cooperative learning group, scored higher than did their counterparts in the lecture class (Giraud, 1997). The study suggested that cooperative learning induces higher achievement in statistics courses than do lecture methods of instruction for all students. Most importantly, students who were least prepared for statistics benefited the most from cooperative learning.

Over the course of several semesters, Keeler and Steinhorst (1995) changed from delivering course content using traditional lecturing to using a cooperative learning approach. The purpose of their study was to examine the relationship between grade achievement and retention as related to cooperatively and traditionally structured courses. Researchers cited the following three reasons for experimenting with cooperative learning: (a) to increase completion rates, (b) to improve students' performance on papers and tests, and (c) to improve students' attitudes towards the course and subject. Data comprised final grade distributions, student retention in class, and a questionnaire that queried students' attitudes towards group activities. Keeler and Steinhorst found that a greater number of students successfully completed the courses using the cooperative learning model, and students achieved higher final scores in the two experimental sections than in the comparison course section. Moreover, students' attitudes towards the cooperative group learning experience were positive. Interestingly, better scores on the first two exams by the traditional semester students were offset by a comprehensive final exam that was 20 points lower than the first two exams. Consequently, the class mean cumulative score in the traditional section was half a letter grade lower than in the cooperative learning classes.

Purpose of the Present Study

In summation, cooperative learning activities appear to benefit students ranging from elementary to baccalaureate educational levels. Johnson et al. (1991a) suggest that the benefits are the result of bi-directional relations among achievement, quality of interpersonal relationships, and psychological health--with each impacting the others. Unfortunately, scant research exists at the graduate level (Slavin, 1991b). Because research methodology courses are taken by the vast majority of the graduate students, and because many students perform at lower levels in these courses than in any other course in their programs of study (Onwuegbuzie, 1997), the purpose of this study was to investigate the effects of cooperative learning on levels of achievement in these classes. An extensive review of literature revealed no study comparing the effects of cooperative learning and traditional structured teaching techniques in research methodology classes. The aforementioned findings that cooperative learning tends to increase levels of achievement in statistics classes, coupled with the fact that research methodology classes typically contain elements of statistics (Mundfrom, Shaw, Thomas, Young, & Moore, 1998), led to the hypothesis that students assigned to cooperative learning groups in research methodology courses have higher levels of achievement than do their counterparts who are not assigned to groups.

A secondary purpose was to determine the effects of cooperative learning on students' attitudes towards group activities and overall learning in research methodology courses. It was expected that the majority of students assigned to cooperative learning groups would find this instructional method more helpful than obtaining information solely via a lecture format. In particular, it was hypothesized that the majority of students in the cooperative learning groups would report positive attitudes towards this method of instruction. It was hoped that the results of this study would assist in finding optimal ways to enhance student learning in research methodology courses.

Method

Participants

Demographics. Participants comprised 193 students, enrolled in graduate-level educational research methodology courses within a three-year period at a mid-southern university. Eighty-one students were enrolled in sections in which cooperative learning groups were formed to undertake the major course requirements (i.e., written critical evaluation of published research reports and preparation of research proposals). The remainder ($n = 112$) were enrolled in sections in which all assignments were undertaken and graded on an individual basis. The same instructor taught all sections.

With respect to the cooperative learning groups, the majority of participants was female (79.8%), ranging in age from 22 to 55 ($M = 32.4$, $SD = 8.5$), with a mean grade point average of 3.67 ($SD = 0.39$). The racial composition was 84.0% Caucasian-American and 16.0% African-American. With regard to the control group sample, most of the participants were female (82.5%) and Caucasian-American (87.5%), ranging in age from 23 to 60 ($M = 31.1$, $SD = 8.6$), with a mean grade point average of 3.63 ($SD = 0.39$). No significance difference ($t = 1.28$, $p > .05$, effect size = 0.10) in mean grade point average was found between students in the cooperative learning classes and those in the individualized classes.

Procedure

Course requirements. According to the university graduate handbook, the course involved the "application of scientific method to educational research, including nature of research problems in education, theory of research, experimental design, techniques in data gathering, the interpretation of results, research reporting, and bibliographical techniques." For each 16-week semester, classes were held for three hours, once per week. The main requirement of the course was the completion of a research proposal. The objective of the proposal was to prepare students throughly to be able to write proposals for dissertations and for seeking external funding. As such, the research proposals provided authentic assessment. Specifically, the research proposal, which could represent either quantitative or qualitative research on a topic of their choice, had to comprise a title, introduction section, review of the related literature, methodology section, analysis section, bibliography, and appendix section including a biography of proposal writer, time-table, budget, consent form(s), and author-designed instrument(s). Research proposals had to be unique and realistic, have educational significance, and extend the knowledge base. Students were expected to type their proposals, following guidelines specified by the American Psychological Association (1994). Students' writing style (e.g., grammar, punctuation, clarity, and application of the American Psychological Association (1994) criteria) also was assessed. All proposals had to include an in-depth review of the literature, and thus extensive library usage was required. Indeed, although many research methodology instructors appear to require what could be conceptualized as a *mini-proposal*, the research proposal in this course was required to be extremely comprehensive. Historically, over the years, research proposals in this course typically ranged from 25 to 40 pages, with the literature review section usually ranging from 5 to 15 pages. Students in both treatment conditions were encouraged to immerse themselves with their research proposals from the first class meeting. Moreover, students/groups were required to formulate their research questions by the second class meeting and to start obtaining literature sources by the third class meeting.

The second major course requirement involved a written critical evaluation of a published research report (article critique). The major goal of the article critique was to allow students to practice evaluating published research articles utilizing principles of the scientific method. In order to prevent student procrastination, students were required to select several potential articles to critique and to bring them to the second class meeting for advice from the instructor as to their appropriateness. Furthermore, students were required to make their final selection as to which article to critique by the third week of the semester.

Because students in both treatment groups typically had various levels of experience using the library, a one-hour library orientation always was provided for them at the second class meeting. In this orientation, the library coordinator demonstrated how to conduct extensive library searches, showing them how to use several electronic databases (e.g., ERIC, PsycLIT, MEDLINE, CINAHL), as well as familiarizing students with the location of various sources (e.g., periodicals) and interfaces. In short, this orientation involved course-integrated instruction, live interactive demonstrations, and hands-on practice exercises. Student feedback indicated that most students found this orientation to be extremely useful (Onwuegbuzie, 1997).

Due to the comprehensiveness of the article critique and the research proposal, the instructor attempted to make himself as available as possible to all students outside class time and office hours, encouraging them to contact him at his home between 10 a.m. and 10 p.m., on any day of the week (including weekends and holidays), if they had any questions about the assignments. Many students in both experimental conditions took advantage of this opportunity.

Individualized learning group. The first part of each class period typically consisted of a review of the material presented in the previous period. Problems which were assigned were discussed and questions answered. The middle portion of each class lesson generally involved presenting new material. The instructor's style was primarily lecture interspersed with students' questions. He used a didactic approach which precluded opportunities to engage in cooperative learning within the lecture period. Visual aids used primarily were overhead projectors and a chalk board. The last part of the class period tended to contain an overview of the reading, a presentation of the follow-up tasks assigned for the next period, and a brief overview of the material to be covered in future sessions. All students were provided with a complete set of the instructor's lecture notes at the beginning of the course. These notes comprised a summary of the textbook and a synthesis of other relevant sources.

Cooperative learning group. On the first day of class, students, in turn, were asked to introduce themselves to the whole class, disclosing their major, educational aspirations, profession, and interests. Following these introductions students were asked to form groups comprising 3-4 students. Students were encouraged to choose group members based on major, profession, and proximity to each other's homes. A few groups involved pairs. These pairs were formed when the class size represented a prime number. The cooperative learning group that was utilized involved the use of base groups (Smith et al., 1992). The aim of these base groups was to promote stable membership whose foremost responsibility was to provide each student the support, encouragement, and assistance needed to understand the material presented by the instructor and in the readings, with a view to (a) completing the group assignments successfully and (b) preparing students for the in-class examinations. Students were encouraged to stay together during the entire course. Although they were allowed to change groups if any conflicts or unresolvable problems arose among group members, no student requested such a change. Students were asked to exchange telephone numbers and e-mail addresses and information about their schedules so that they could meet outside class. Each base group undertook one research proposal and one article critique.

The instructor informed students of the following basic group skills: every group member should participate as equally as possible, or at least according to their strengths, students should respect the opinions of all group members, no students should dominate group discussions, and every student should be aware of all tasks undertaken by group members and be prepared to provide constructive criticism. Students were not assigned specific group roles; however, they were presented with different models for the division of labor (e.g., each student writing a section of the research proposal and article critique; each student individually undertaking all sections of these assignments and then comparing their work with all other group members with a view to merging).

As in the individualized groups, the first part of each class period typically consisted of a review of the material presented in the previous session and the middle portion of each class lesson generally involved the presentation of new material. All students were provided with a complete set of the instructors' lecture notes at the beginning of the course. However, instead of a lecture-based review of the material, as in the individualized groups, in the cooperative treatment condition, each base group reviewed the material that was presented earlier by the instructor. During this phase, students rearranged desk-chairs into groups within the classroom. While students worked in groups, the instructor observed, answered questions posed by students, and informed the class of any insights gained from circulating among the groups. As time permitted, students in the cooperative groups also were given class time towards the end of the period to discuss their research proposals and/or their article critiques.

Instruments

A scoring rubric was used to evaluate proposals and article critiques, with detailed feedback provided. Students in the control group received individual scores, on a 100-point scale, for their research proposals and article critiques. Students in the cooperative learning groups were given group scores for these assignments. Conceptual knowledge, which involved students' knowledge of research concepts, methodologies, and applications, was measured individually in both sets of classes via comprehensive written midterm and final examinations. The examination form consisted of open-ended questions, involving items which required knowledge of the research process. All of the items pertained to content from the first half of the course and were chosen from the instructor's item bank to ensure that the examination was typical of past examinations given by this instructor. The final examination also was constructed by the course instructor and paralleled the format of the midterm examination, yet covered the complete course content. Both the midterm and the final examination were administered under untimed conditions and were scored on a 100-point scale by the instructor, using a key that specified the number of points awarded for both correct and partial-credit answers. Finally, students in the cooperative learning were asked to keep reflexive journals. Journal entries were recorded by students on a regular basis.

Results

Analysis of Achievement Data

A *split-plot* analysis of variance (ANOVA) was undertaken using treatment group as the between-subjects factor and examination time as the within-subjects factor (Maxwell & Delaney, 1990). The Shapiro-Wilk test (Shapiro & Wilk, 1965; Shapiro, Wilk, & Chen, 1968) did not indicate that the distribution of the midterm examination ($W = .97$, $p > .05$) or the final examination ($W = .97$, $p > .05$) scores was non-normal. Also, because there were only two levels of the within-subjects factor (i.e., the midterm and final examination), the sphericity assumption (i.e., the homogeneity of treatment-difference variances assumption) was met by default (Maxwell & Delaney, 1990). (It should be noted that a multivariate analysis could have been undertaken to analyze the data rather than the univariate split-plot ANOVA. However, when there are only two levels of the within-subjects factor, as is the case here, the two approaches are identical. See for example, Maxwell & Delaney (1990).

The split-plot ANOVA revealed a treatment group x examination time interaction ($F_{1, 191} = 5.80$, $p = 0.017$; $\hat{\omega}^2 = .05$). Both main effects also were statistically significant, namely, treatment group ($F_{1, 191} = 4.74$, $p < 0.031$; $\hat{\omega}^2 = .02$) and examination time ($F_{1, 191} = 18.79$, $p < 0.001$; $\hat{\omega}^2 = .08$). Follow-up tests of simple effects indicated that students enrolled in cooperative learning sections of the course had statistically significantly ($t = 3.01$, $p < .01$) lower levels of performance ($M = 76.7\%$, $SD = 12.0$) than did their counterparts ($M = 82.1\%$, $SD = 10.7$) at the midpoint of the course, as measured by the midterm examination. The effect size (ES) corresponding to this difference (i.e., mean differences divided by the pooled standard deviation, which is known as Cohen's *d*) was .48, which, using Cohen's (1988) criteria, suggests a moderate effect. Although students in the cooperative learning groups still had lower levels of performance ($M = 82.2\%$, $SD = 9.7$) with respect to the final examination than did students in individualized learning sections ($M = 84.9\%$, $SD = 11.7$), this difference was not statistically significantly different ($t = 1.68$, $p > .05$). However, the corresponding effect size (i.e., 0.25), suggests that this difference may be non-trivial. The simple effects indicated that an ordinal interaction was apparent, with individualized learning groups obtaining higher scores on both in-class examinations, with the difference between the two groups being statistically significantly higher only at the midterm stage. Finally, no overall difference ($t = -1.21$, $p > .05$, ES = 0.16) in course average was found between the cooperative learning groups ($M = 83.8\%$, $SD = 6.7$) and the individualized learning groups ($M = 82.4\%$, $SD = 9.8$).

Interestingly, although scores obtained for the research proposals did not discriminate the cooperative learning ($M = 82.5$, $SD = 15.3$) and the individualized learning ($M = 81.8$, $SD = 11.3$) groups ($t = 0.3$, $p > .05$, ES = 0.04), the article critiques completed by students in the cooperative learning groups ($M = 85.3$, $SD = 8.4$) obtained statistically significantly higher scores ($t = 2.3$, $p < .05$, ES = 0.32) than did those undertaken individually ($M = 81.5$, $SD = 14.5$), with a moderate effect size.

Analysis of Qualitative Data

A phenomenological mode of inquiry (*inductive, generative, and constructive*) was used to analyze the reflexive journals (Goetz & Lecompte, 1984, p. 54). In order to determine the percentage of students who responded positively to their cooperative learning experience, these data were *unitized*, that is, units of information served as the basis for defining a significant statement (Glaser & Strauss, 1967). Each unit corresponded to a significant statement in each journal, which provided evidence of student attitudes towards their cooperative learning groups (Lincoln & Guba, 1985). Each significant statement was categorized either as positive or negative. Accordingly, a journal was characterized as being indicative of a positive overall attitude towards cooperative learning if at least two-thirds of the significant statements were positive. Similarly, a journal was characterized as being indicative of a negative overall attitude towards cooperative learning if at least two-thirds of the significant statements were negative. Finally, a journal was characterized as being indicative of an ambivalent overall attitude towards cooperative learning if between one-third and two-thirds of the significant statements were negative/positive. Based on these classifications, 70.4% (n = 57) of the participants were determined to have positive overall attitudes towards their cooperative learning experience, 19.8% (n = 16) of the participants were determined to have negative overall attitudes, and 9.9% (n = 8) were regarded as being ambivalent.

The *method of constant comparison* (Glaser & Strauss, 1967, p. 105) was utilized in order to categorize units which appeared similar in content. Each category represented a distinct theme. This method of analysis revealed a number of themes relating to students' attitudes towards their cooperative learning experience. Each emergent theme was indicative of either a positive or negative cooperative learning experience. Each of these themes will now be discussed with examples from the database used as illustrations. As needed, pseudonyms are used to maintain confidentiality.

Positive Themes

The most consistent positive response was that the use of cooperative learning groups allowed students to interact with other group members and to share ideas. Examples of this positive interdependence include:

> I thought cooperative learning was an excellent experience. I liked getting to know others in my field and working with them. Even though we were at different levels of understanding of the assignments and writing abilities, modifications had to be made to part of the critique to bring it up to standard. I would not hesitate to do this again--this is how things work in real life--collaboration with other professionals.
>
> I liked proofing each other's work to produce a quality product and will do this again with other papers. I liked utilizing each other's strengths and learning from each other. Two heads are better than one.

One student declared: "I have really enjoyed working in the cooperative group in this class. I feel it has given me someone to work with as well as share ideas . . . ." Another student noted that the cooperative learning groups "help everyone understand the different areas of the critique by having someone to talk to about it." Evidence also was provided that the cooperative learning groups prevented some of the weaker students from losing focus: "Being in the group helped me to see other viewpoints and also helped stay focused on the right track because sometimes I went 'blind'."

Some students believed that being in a group that was homogeneous with respect to area of study was advantageous: One group member stated: "It helped that each were majoring in the same area." Similarly, another student reported that "Since we all had a common interest (early intervention special education) we worked well together." The ability to provide or to receive peer-tutoring (i.e., scaffolding) was another positive outcome cited by some participants in the cooperative learning groups. The following statement exemplified scaffolding: "Cooperative learning allows students to learn from their peers. This may be helpful when the teacher is unable to explain something well enough for everyone in the class to understand." Another student observed that: "Everyone has a different learning style and this [cooperative learning] would give people that need help a chance to work with someone who could help them." Additionally, one group member concluded: "I learned more from my classmates than by doing it [the research projects] all by myself."

A few students found that cooperative learning helped to increase their levels of self-esteem and self-efficacy. For example, one student revealed the following:

> I am not an outgoing person. I do not always feel confident in myself to do some assignments. I feel the assignments would have been hard for me to do myself. I have low self-esteem. I feel that working in cooperative learning groups helps to build up your self-esteem.

The dimension of personal responsibility was evidenced when one student disclosed: "I feel that I put more effort into the paper than I would have if it had been just for myself because I knew that other people's grades were going to be affected also." Even students who tended to procrastinate appeared to benefit from cooperative learning, as illustrated by the following journal entry:

> I am a procrastinator and having a group to help me get my part done on time made it easier to finish with less stress and worry. I like to put things off until the last minute, and then rush through. I found it helpful to get things done ahead for a change.

Many participants believed that cooperative learning helped to reduce their levels of anxiety. Compelling accounts of how cooperative learning reduced anxiety levels are reported below:

> I enjoyed working with my co-workers. We got to know each other better by working together--that made the class more 'comfortable.' It seemed like when one was anxious, the others were encouraging. We supported each other.

Another student admitted that "I really enjoyed working on this as a cooperative learning group. It helped to reduce my level of anxiety." Yet another student reasoned that "By working in groups, the stress and anxiety was shared and therefore somewhat lessened."

Some of the groups provided ample evidence of positive interdependence and face-to-face promotive interaction. The following description provides an example in which these features appeared to be maximized:

> At the very beginning we were all concerned that our varied educational backgrounds would cause difficulty. This was not a problem. We each brought different strengths into the group so we were made stronger as a whole.

> I feel the members of our group got along together wonderfully. We worked on everything together and everyone seemed willing to cooperate. I can't think of any instances when we had a problem.

Notably, social cohesiveness enhanced the quality of interpersonal relationships among students, with some group members stating that this method of instruction helped to improve the classroom climate by increasing the incidence of positive, cooperative interactions both within and outside the classroom. Indeed, many students reported enjoying sitting close to their group members in the classroom. Perhaps even more compelling was the fact that cooperative learning appeared to have the potential to create permanent bonds among some of the students: "With this style of work, we have made new friends in the process."

With respect to problem solving, cooperative learning appeared to be effective for some groups. In many cases, when students worked on a problem, such as arriving at a final draft of the research proposal, the problem was solved redundantly, which enabled students to check each other's problem-solving procedures and mistakes during the process. As noted by one student: "It was nice having other people to check my work. They caught some mistakes that I had missed. It was a great learning experience!" Another student declared: "I liked proofing each others' work to produce a quality product and will do this again with other papers [in other classes]." Yet another student revealed: "Although we each were assigned a section of the projects, we got together after completing the sections and re-worked them as a group." A few individuals recognized that the groups' attempt to maximize the quality of their projects increased the length of time spent on solving tasks, as evidenced by the following:

> I believe that our group effort probably took longer to write the critique than if we had written it individually. However, I feel that we probably discussed the information in greater detail as we critiqued the article.

An important finding was that the groups who appeared to be the most functional with respect to the distribution and completion of tasks also appeared to have the most metacognitive awareness of the cooperative learning process and to exhibit more self-regulatory behaviors than did members of less functional groups. In addition, functional groups tended to assign roles to each group member and to distribute the workload as equally as possible. For example, a member of an extremely functional group noted the following:

> We promoted each other's learning by helping, sharing and encouraging. We could explain, discuss, what we knew to each other. We talked through each aspect of the assignments. In cooperative learning groups you need a leader, decision maker, trust builder, communicator and working relationships among members. Our group had all of these. Sarah and I contributed in the projects being an encourager, praiser, to reinforce members' contributions. Terri and Jacquie were the recorders to write down the groups' decisions and edit the group's reports. Laura volunteered to be the typist for the group. Our group stayed on task and maintained a good working relationship.

Many of the inservice teachers related their cooperative learning experience to their own classrooms. For example, one teacher admitted: "I have gained new insights into cooperative learning and team work that I can take back and use in my classroom (i.e., the necessity of good communication and conflict management)." Another compelling statement posited by a teacher was:

> I also think that it is a good experience for adults (mainly teachers) to get another experience with cooperative learning. It gives us a chance to see what our kids go through again since most [teachers] use cooperative learning in their classes. But haven't had a practical experience with it in a while.

Negative Themes

Although the negative themes which emerged were less frequent in occurrence than the aforementioned positive themes, many of them were, nevertheless, noteworthy. By far the most cited criticism of their cooperative learning experience centered around the time constraints. Many students noted how difficult it was for them to meet group members outside the classroom--especially students who lived the furthest from their group members. One student noted that "the only negative experience we had was finding a convenient time for all to meet." Another student stated: "The only drawback I have is the time factor. It required some schedule balancing to find time to get together and work outside the class." One student declared that her inability to meet regularly with her group members would disrupt the group:

> I don't think it [cooperative learning] worked well because our group had a hard time finding a time to meet--we all have jobs, schedules, families--it was very difficult. I, especially, felt like I had abandoned the group because of things going on in my personal life. It would have been more convenient to work alone.

Closely related to the issue of time was the issue of location. One student declared: "I don't like cooperative learning. It was difficult to meet in groups because of location." Another student wrote: "Our main disadvantage was having to travel in order to work together." This was again echoed by the following attitude: "I do not like cooperative learning groups because . . . when your group *all* live in a different location, it is very hard to *all* get together to work on the project." In an attempt to resolve some of the scheduling and logistical difficulties, two group members recommended that "more class time be devoted to working in groups." Additionally, two individuals suggested that reducing the group size to two students would alleviate some of the time and location constraints. One of these students wrote the following:

> I am not sure this setting is very effective for cooperative learning. Too many different occupations and professions involved. As for me personally, I would have rather just, let us get in pairs--probably could have gotten better working conditions--due to time frame-professions, etc.

The other student revealed:

> I usually have no problem working in groups, however, this time it was difficult. With everyone in my group working and going to school, it was hard to find time to get together. Also, all three of us lived in three different cities. I think for smaller projects, group work is terrific, but writing papers of this length should be assigned to one person, or two at the most.

One student complained that her group often arranged to meet at inconvenient times for her: "At the time my group members wished to meet to work on their critique, I should have been and needed to study for my mid-term." However, not all students were negative about having to schedule meetings outside the class sessions. In particular, one student revealed: "I admit, I was skeptical at first. However, as time went on, I saw that we could accommodate everyone's schedules to find time to meet."

Although, as evidenced earlier, many students experienced increases in their motivation levels in an effort to help maximize their group's grade, some students appeared to their group members to remain unmotivated. One student thought that she "was put in a group that wasn't motivated." Apparently, this group took a long time to start their assignments. Thus, an unfortunate outcome of cooperative learning enterprise was that a few students used the opportunity to coat-tail. One individual made the following observation:

> I believe that cooperative learning in research methodology courses is really an effective way for instruction. However, I realize that even in graduate school there are those who will not take the responsibility of "sharing the load." When this happens it tends to bring down work quality of the entire group.

A student who attained one of the highest overall averages in the course noted that "It was hard for everyone to have an equal share since some people were reluctant to contribute to the group's discussions." Similarly, another high-achieving student complained that "two group members gave no feedback or assistance in the proof-reading and correcting of errors on the finished paper." Unfortunately, these students did not disclose which group members they regarded as coat-tailers. However, one low-achieving student admitted to coat-tailing:

> I have to be honest about the work this semester. I probably did not do but 40% of the work but the circumstances did not allow me or my partner to change that fact. He did not mind and he knew that if the shoe had been on the other foot I would not have minded either. When you are partners you can't always split the work 50/50 but you can put forth your best effort and do what is required. I appreciate the chance for the opportunity to work with a group, thank you.

Although, as documented above, many students reported that their cooperative group experience helped to *reduce* their levels of anxiety, one student disclosed that cooperative learning *increased* their anxiety levels: "Our groups worked extremely well together. We were all concerned that our work would negatively affect the others' grades. That fact increased my anxiety level." Trying to reach a consensus also was a source of anxiety:

> While our group was small (only two), we had some difficulties agreeing on what we wanted to say. I think larger groups would experience this problem even more. My partner and I have similar learning styles and work habits so that was not a problem, but several people would be difficult to work with.

One student astutely noted that "cooperative groups work well if people respect individual's ideas." Unfortunately, it was clear that some of the high-achieving students did not respect their lower-achieving group members. For example, one student declared: "I was disappointed with one peer's work and making her redo it didn't seem to help [its quality], so we had to redo it [for her]." Another high-achieving student revealed the following:

> In general, I dislike cooperative learning intensely. I don't like to be involved in group work because I am very independent and picky about work quality. . . . The difficulties I had were because I found it somewhat difficult to clarify other people's thoughts when they didn't seem straightforward, and I was displeased with the literature review section but did not have time to redo it since I got it "late." I don't like to "share" grades.

This account suggests an intolerance for what they regarded to be sub-standard work. This intolerance, coupled with a lack of trust of some group members' ability to contribute effectively to the assignment, led to some high-achieving students undertaking a disproportionate amount of the workload. For example, one high achiever stated: "I do not like cooperative learning. I ended up doing a larger share of the work because I wanted the highest quality of work. In my case cooperative learning is more work than completing a project on my own." Another high performing student admitted: "I may have appeared overbearing to the group because I wanted to have everything checked and re-checked. The group was gracious about this." Yet another high-achieving student revealed:

> I was responsible for the summary, the majority of the introduction, and for putting it all together and checking APA style. I do feel like I did most of the work in the group, but also I feel like I took on that role myself because that is what I needed to do before I put my name on the paper. I'm too paranoid to turn something like this over to others, so I accept that as one of my hang-ups in regard to cooperative learning.

Perhaps the most compelling finding that emerged from the phenomenological analysis centered around the group structures. In particular, the level of homogeneity appeared to play a role in determining how functional each group was. It appeared that groups which were too heterogeneous tended not to function as well as did homogeneous groups. Age was one factor which appeared to be a barrier in some cooperative learning groups. For example, one student noted that "because I am older and tend to be more set in my ways, I found it frustrating to agree on ideas."

By far the most important grouping factor was ability with respect to the research methods course. Interestingly, it appeared that the more homogeneous with respect to ability the group was, the more positive features they displayed. These features included positive interdependence, social cohesion, and a willingness to promote each other's learning and to hold each other personally and individually responsible for doing her/his fair share of the work. For example, the group which clearly was the most functional comprised members who all ended up with either average or below average overall achievement levels. This group did not complete the best projects, but they were the most positive about their cooperative learning experience--with all of them stating that they expected to remain friends after the semester ended, even though they did not know each other prior to the course.

Conversely, the group containing students with the widest range of ability appeared to encounter the most problems. Whereas the weakest member of this 5-person group stated that she liked the cooperative experience, and the next weakest member expressed concern that the stronger members of the group dominated the projects, the highest-achieving member was extremely critical of the weaker members, as follows:

> I have an overwhelming dislike of this method of teaching. First, let me explain that I am strongly motivated by my desire to maintain a 4.0 in my graduate studies. Also, I have a very difficult time allowing the work of someone else to determine the grade I obtain, therefore I tend to take over and do all the work myself. In this instance, I had another compatriot who also is driven by a desire to make A's and is also driven for excellence in her work. I am not sure whether it was our overbearing qualities that laid most of the work on our shoulders or our fear that if we didn't do it it wouldn't get done. The third member of our group tried extremely hard to participate in all of the assignments but she was relatively clueless about how to do it. We tried to explain as best we could, but under the time constraints I didn't know that we did a very good job. I know that I was disappointed in the lack of effort on the part of one member in particular who made no attempt to work on either project. She was not dependable, promised to do things then did not, made dates to be there but didn't show up . . . you get the picture.
>
> This experience has forced me to reevaluate the use of group teaching in my classroom. I now have empathy for those motivated students whose desire to succeed means they will do ALL the work rather than fail, even if it means other students get to reap the A's they produce. Fortunately, my education is more important than my grade, and I know that when I leave this class, my ability to write, to present, and carry out quality research will be there. This is not true for one or two others in the class.

Although this student was obviously angry about the contributions or lack thereof of the "weakest" members of her group, she recognized that her workload, although high, would hold her in good stead in the future.

Discussion

The purpose of this study was to investigate the effects of cooperative learning on levels of achievement and attitudes towards group activities and overall learning in research methodology courses. Findings revealed a statistically significant interaction between treatment group and examination time. More specifically, students in cooperative learning groups had significantly lower performance levels than did their counterparts at the midpoint of the course, as measured by the midterm examination. Indeed, the effect size pertaining to this difference was moderate. Moreover, using the predefined grading scale of the course instructor, the 5.4 point difference between the groups represents one-half of a letter-grade.

No statistically significant difference was found between the groups with respect to the final examination, although this difference may have been non-trivial, in favor of the individualized learning group. Because the two groups did not differ with respect to grade point average, it is unlikely that past academic achievement was a major factor in explaining these findings. Similarly, unlike most studies in the area of cooperative learning, it is unlikely that teacher variables played a major role, because the same instructor taught all groups in the study.

The fact that the cooperative learning groups had lower levels of achievement at the midpoint of the course, but that by the end of the course this differential decreased significantly (as indicated by the statistically significant interaction effect), suggests that cognitive outcomes improved over time. This, in turn, suggests that collaborative techniques may need time for their effects on achievement levels to be realized. Indeed, the result relating to the midterm deficits is consistent with the qualitative finding that some groups were slow in starting their projects. Moreover, the finding of no significant difference in the final examination scores is consistent with Courtney, Courtney, and Nicholson (1992), who found no differences in statistics achievement between graduate students who were taught using a cooperative learning method and those who were taught using a traditional method. Nevertheless, the fact that the gap in achievement levels between students in the two groups substantially narrowed by the end of course may be attributable to scaffolding, that is, to more able students helping the less-prepared students to understand the material presented.

Interestingly, no difference in overall course average was found between students in the two groups. This may be attributable to the fact that, although scores obtained for the research proposals did not discriminate the cooperative learning and the individualized learning groups, the article critiques completed by students in the cooperative learning groups obtained statistically significantly higher scores than did those undertaken individually. Clearly, the article critique assignment and, to a smaller extent, the research proposal, helped to increase the course average of students in the cooperative learning groups relative to their counterparts, culminating in no overall difference in course grades between the two groups. Bearing in mind the comprehensiveness of the group projects, it is not surprising that the cooperative learning students performed at higher levels for at least one of them.

It is likely that the higher midterm examination scores obtained by students in the individualized learning group arose because they did, or at least were expected by the instructor to do, as much work on the two assignments (i.e., the article critique and research proposal) as were the entire group in the cooperative setting. That is, the additional effort required by the individualized group may have accounted for at least some of the higher achievement experienced by this group. Although it cannot be assumed that the extra effort per capita translates into a better product (indeed, this was not the case for research proposals), it is possible that working alone on the article critique and research proposal promoted their understanding of the research process to a greater extent than for those in the cooperative learning groups--particularly the coat-tailers.

In any case, the finding that the two groups did not differ in overall achievement in the research methods course supports Davidson's (1985) contention that conditions of cooperative learning groups such as group incentives which are required for success at the public school level may not be effective at the college level. Indeed, the similarity of overall performance levels found in the present study suggests that cooperative learning may not lead to grade inflation, provided that there is individual accountability. It appears that the in-class examinations in this course helped to ensure individual accountability.

Although the hypothesis relating to achievement was not supported, the qualitative analyses indicated that the vast majority of students like cooperative learning. Indeed, the 70% of cooperative learning students who responded positively suggests a large effect size. In particular, the phenomenological analysis revealed that most students in the cooperative learning groups experienced the predicted shifts in motivation, persistence, self-esteem, self-efficacy, anxiety, social cohesion, problem solving adeptness, and metacognitive awareness. Because these variables represent positive outcomes of cooperative learning for many subject areas and at most age levels (Johnson et al., 1981), and because many of these variables have been found to be related to achievement in research methodology courses (Onwuegbuzie, 1997; Onwuegbuzie & Daley, 1996; Onwuegbuzie & Seaman, 1995; Onwuegbuzie, Slate, Paterson, Watson, & Schwartz, 2000), it is perhaps surprising that students in the cooperative learning group did not experience higher levels of achievement than did their counterparts. Thus, future research should investigate the reliability of the lack of difference in achievement levels found in the present study.

The fact that students appear to like cooperative learning techniques despite not experiencing increases in their levels of performance, may be related to the affective benefits of this form of learning. Numerous positive statements suggest that, for some students, the non-cognitive outcomes may be as important as subject matter achievement. This finding is congruent with earlier conclusions that cooperation promotes self-esteem, caring relationships, and psychological health (Johnson & Johnson, 1989; Sharan, 1990; Slavin, 1990).

Nevertheless, a few students, particularly the weaker ones, seemed to have liked cooperative learning for reasons which are not compatible with the instructional objectives of this method. That is, these coat-tailers appeared to like cooperative learning because they realized that they do not have to put forth as much effort in order to obtain a passing grade in these courses. These students may then rely on their more able group members to maximize their groups' project grades. The possible inflated project grades on the part of the weaker students may, in turn, have reduced the pressure on these less able students to achieve in the in-class examinations, culminating in reduced levels of motivation to study and, subsequently, lower actual performance levels--especially at the midterm stage. It is also possible that some of the weaker students, especially in heterogeneous groups, were not able to make a large contribution to their groups due to the domination of the workload by their more able counterparts. Such unequal distribution of the workload may have prevented weaker students from taking an active role in the whole research process, thus debilitating their performance levels. Indeed, Cohen (1994) and Slavin (1996) have noted the importance of fairness in work load.

An important limitation of the present investigation is that the results were obtained from a relatively small, non-random, geographically-limited sample of students seeking graduate degrees. Thus, the extent to which the findings generalize to other students enrolled in graduate programs is a question awaiting subsequent research using both quantitative and qualitative analysis techniques, as in the current study. Another threat to external validity stems from the fact that one of the researchers was the instructor of the course. Specifically, this threat, which is termed experimenter/researcher effect (Gay & Airasian, 2000), may have biased the findings of the study to some degree. However, it should be noted that the inter-rater reliability between the two observers with respect to the identification of the emergent themes and the classification of significant statements to these themes was extremely high (i.e., 100%).

Reactive arrangements was another threat to external validity (Gay & Airasian, 2000). In particular, it is possible that the overwhelming positiveness of the journal entries may have been, in part, related to social desirability. Because the journals were turned in to the instructor who was responsible for assigning course grades, it cannot be ruled out that at least some of the students might have avoided making negative statements. However, it is possible that this threat was minimized for the following two reasons: (1) students were guaranteed maximum points for their journals, provided that entries were recorded on a regular basis, and (2) students were informed that their journals would not be read by the instructor until grades had been assigned and submitted to the registrar's office. Furthermore, the majority of students noted at least one negative statement about the cooperative learning process--many of whom cited the time and location constraints. If social desirability had played a large role in determining students' journal entries, it is likely that many more students would have recorded only positive responses.

A threat to internal validity was instrumentation. Due to the open-ended nature of both midterm and final examinations, assessment of internal consistency was not possible. However, it is likely that the use of detailed scoring rubrics helped to increase the reliability and validity of scores yielded by these measures.

A weakness of the research design was the fact that participants in the individualized group were not asked to complete journal entries. Such information could have led to comparisons of attitudes across each group. Interestingly, however, Onwuegbuzie (1997) analyzed journal entries of students who were enrolled in research methodology courses wherein the individualized method was utilized. This researcher noted the particular difficulties that many students had writing their research proposals, as well as the continual high levels of anxiety. A comparison of the journal entries of students in the cooperative learning group in the present investigation and the individualized learning students in Onwuegbuzie's (1997) study revealed that the cooperative learning students made less references to their anxiety levels than did the individualized learners. Indeed, in some instances, when anxiety levels were discussed by the cooperative learning participants, it was to note that the cooperative learning process helped to *reduce* their levels of anxiety.

Nevertheless, future research should compare comments made by both the individualized learning group and the cooperative learning group within the same study. In order to facilitate such a comparison, "high specificity" journal entries would be needed. Such journals would involve more structure than the "low specificity" entries that were required for the present investigation. For example, a semi-structured journal format could be designed, whereby students are asked to respond to a few open-ended questions while recording their journal entries. Alternatively, a structured journal format could be incorporated in which students are requested to respond to specific questions in a definite order--some of which may be closed-ended.

Many researchers (Johnson et al., 1991a, 1991b) contend that (a) positive interdependence, (b) a willingness to promote each other's learning, (c) holding each other personally and individually accountable for her/his fair share of the work, (d) using appropriately the interpersonal and small-group skills needed to maximize group effectiveness, and (e) self-monitoring of how adequately members are working together, are essential for cooperative learning to be beneficial. Unfortunately, one or more of these five elements were not present in some of the groups--especially the most heterogeneous groups. This finding suggests that research methodology instructors who utilize cooperative learning techniques should emphasize to students the importance of maintaining these five elements, and should monitor each group for their presence on a continual basis.

In any case, the findings that high-achieving students in heterogeneous groups typically were the most critical of their group members suggests that an aptitude *x* treatment interaction took place in this study. This possible interaction should be the subject of future investigations.

In summary, most graduate students in this study appear to regard cooperative learning as an effective instructional method in research methodology courses. However, their attitudes appear to be at odds with their test performance. As Johnson (1992) noted, there is an important difference between simply putting students in groups to learn and in structuring cooperation among students. Although the cooperative learning groups in this study were structured, comprising heterogeneous base groups, in which participants stayed together during the entire course, the findings suggest that even greater cooperative structure is needed in the course. The present authors currently are investigating the effects of increased cooperative structure on achievement and attitudes in research methodology courses, as well as the effects of the instructor monitoring group processing, modeling problem-solving skills, providing regular feedback regarding individual and group mastery, and evaluating group effectiveness on a regular basis. It is hoped that such studies will help to determine conditions under which the benefits of cooperative learning are maximized.

References

American Psychological Association (1994). *Publication manual of the American Psychological Association.* Washington, D.C.: American Psychological Association.

Borresen, C.R. (1990). Success in introductory statistics in small groups, *College Teaching, 38,* 26-28.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: John Wiley.

Cohen, E.G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64,* 1-35.

Cooper, J., & Muech, R. (1992). Student involvement in learning: Cooperative learning and college instruction. In A. Goodsell, M. Maher, & V. Tinto (Eds.), *Collaborative learning: A resource book in higher education* (pp.68-74). University Park, PA: National Center on Teaching, Learning, and Assessment.

Courtney, D.P., Courtney, M., & Nicholson, C. (1992, November). *The effect of cooperative learning as an instructional practice at the college level.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN.

Davidson, N. (1985). Small-group learning and teaching in mathematics: A selective review of the research. In R. Slavin, S. Sharan, S. Kagan, R. Hertz-Lazarowitz, C. Webb, & R. Schmuck (Eds.), *Learning to cooperate, cooperate to learn* (pp. 211-130). New York: Plenum Press.

Davidson, N., & Kroll, D. (1991). An overview of research on cooperative learning related to mathematics. *Journal for Research in Mathematics Education*, 22, 362-365.

Deutsch, M. (1949). A theory of cooperative and competition, *Human Relations, 2,* 129-152.

Fraser, S.C., Beaman, A.L., Diener, E., & Kelem, R.T. (1977). Two, three, or four heads are better than one: Modification of college performance by peer monitoring, *Journal of Educational Psychology, 69,* 101-108.

Garfield, J. (1993). Teaching statistics using small-group cooperative learning, *Journal of Statistics Education, 1*(1) [On-Line[. Available: http://jse.stat.ncsu.edu:70/0/jse/v1n1/garfield

Gay, L.R., & Airasian, P.W. (2000). *Educational research: Competencies for analysis and application* (6th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Giraud, G. (1997). Cooperative learning and statistics instruction, *Journal of Statistics Education, 5*(3) [On-Line]. Available: http://jse.stat.ncsu.edu:70/0/jse/v5n3/giraud

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory. Strategies for qualitative research.* Chicago: Aldine.

Goetz, J.P., & Lecompte, M.D. (1984). *Ethnography and the qualitative design in educational research.* New York: Academic Press.

Johnson, D.W. (1992). Cooperative learning: Increasing college faculty instructional productivity. (ERIC Document Reproduction Service No. ED 343 465)

Johnson, D.W. (1993). *Reaching out: Interpersonal effectiveness and self-actualization, 5th ed.* Needham Heights, MA: Allyn and Bacon.

Johnson, D.W., & Johnson, R.T. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Book Co.

Johnson, D.W., & Johnson, R.T. (1991). *Joining together: Group theory and group skills.* Boston: Allyn & Bacon.

Johnson, D.W., & Johnson, R.T., & Holubec, E.J. (1991). *Cooperation in the classroom. rev. ed.* Edina, MN: Interaction Book Co.

Johnson, D.W., Johnson, R.T., & Smith, K.A. (1991a). *Increasing college faculty instructional productivity.* Washington, DC: The George Washington University. (ASHE-ERIC Higher Education Report No.4).

Johnson, D.W., Johnson, R.T., & Smith K.A. (1991b). *Active learning: Cooperation in the college classroom.* Edina, MN: Interaction Book Co.

Johnson, D.W., & Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). The effects of cooperative and individualized instruction on student attitudes and achievement: A meta-analysis, *Psychological Bulletin,* 89, 47-62.

Jones, L. (1991). Using cooperative learning to teach statistics, *Research Report* 91-2. L.L. Thurstone Psychometric Laboratory, University of North Carolina.

Keeler, C.M., & Steinhorst, R.K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field, *Journal of Statistics 3*(2). [On-Line]. Available by e-mail: archive@jse.stat.bcsu.edu Message: send jse/v3n2/keller

Lewin, K. (1935). *A dynamic theory of personality.* New York: McGraw Hill.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Maxwell, S.E., & Delaney, H.D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

Mundfrom, D.J., Shaw, D.G., Thomas, A., Young, S., & Moore, A.D. (1998, April). *Introductory graduate research courses: An examination of the knowledge base.* Paper presented at the annual meeting of the American Education Research Association (AERA) San Diego, CA.

National Council of Teachers in Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: The Council.

National Council of Teachers in Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: The Council.

National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academy Press.

Newman, F.M., & Thompson, J. (1987). *Effects of cooperative learning in achievement in secondary schools: A summary of research*. Madison WI: National Center on Effective Secondary Schools.

Onwuegbuzie, A.J. (1997). Writing a research proposal: The role of library anxiety, statistics anxiety, and composition anxiety. *Library & Information Science Research, 19*, 5-33.

Onwuegbuzie, A.J., & Daley, C.E. (1996). The relative contributions of examination-taking coping strategies and study coping strategies on test anxiety: A concurrent analysis. *Cognitive Therapy and Research, 20*, 287-303.

Onwuegbuzie, A.J., & Seaman, M. (1995). The effect of time and anxiety on statistics achievement. *Journal of Experimental Psychology, 63*, 115-124.

Onwuegbuzie, A.J., Slate, J.R., Paterson, F., Watson, M., & Schwartz, R. (2000). Factors associated with underachievement in educational research courses. *RESEARCH IN THE SCHOOLS, 7*(1), 53-65.

Purdom, D.S., & Kromey, J.D. (1992, April). *A comparison of different instructor intervention strategies in cooperative learning groups at the college level*. Paper presented at the annual meeting of the American Education Research Association, San Francisco, CA.

Qin, Z., Johnson, D.W., & Johnson, R.T. (1995). Cooperative versus competitive efforts in problem solving. *Review of Educational Research, 65*, 129-143.

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test, for normality and complete samples. *Biometrika, 52*, 592-611.

Shapiro, S.S., Wilk, M.B., & Chen, H.J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association, 63*, 1343-1372.

Sharan, S. (1990). *Cooperative learning: Theory and research*. New York: Praeger Press.

Shaughnessy, J.M. (1977). Misconception of probability: An experiment with a small-group activity-based model building approach to introductory probability at the college level, *Educational Studies in Mathematics*, 8, 285-316.

Slavin, R.E. (1989). *Research on cooperative learning: Consensus and controversy*. Educational Leadership, 47, 52-54.

Slavin, R.E. (1990). *Cooperative learning: Theory, research, and practice*. Englewood Cliffs, NJ: Prentice-Hall.

Slavin, R.E. (1991a). Research on cooperative learning: Consensus and controversy. In R.S. Brandt (Ed.), *Cooperative Learning and the Collaborative School: Readings from Educational Leadership* (pp.95-97).

Slavin, R.E. (1991b). Synthesis of research on cooperative learning. *Educational Leadership, 48*, 71-82.

Slavin, R.E. (1992). Research on cooperative learning: Consensus and controversy. In A. Goodsell, M. Maher, & V. Tinto (Eds). *Collaborative learning: A sourcebook for higher education* (pp. 97-99). University Park, PA: National Center on Postsecondary Teaching, Learning, and Assessment.

Slavin, R.E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*, 43-69.

Smith, K., Johnson, D.W., & Johnson, R.T. (1992). Cooperative learning and positive change in higher education. In A. Goodsell, M. Maher, & V. Tinto (Eds.). *Collaborative learning: A sourcebook for higher education* (pp. 34-36). University Park, PA: National Center on Postsecondary Teaching, Learning and Assessment.

Smith B., & MacGregor, J.T. (1992). What is collaborative learning? In A. Goodsell, M. Maher, V. Tinto (Eds.). *Collaborative learning: A sourcebook for higher education* (pp. 34-36). University Park, PA: National Center on Postsecondary Teaching, Learning, and Assessment.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 1, 77-82

Likert Survey Primacy Effect in the Absence or Presence of Negatively-Worded Items

J. Jackson Barnette
The University of Iowa

A twenty-item survey was designed in four forms with response set direction as: Strongly Disagree (SD) to Strongly Agree (SA) and SA to SD crossed with the absence or presence of negatively-worded stems. The primary research question related to finding a primacy effect when comparing the response direction formats. Surveys were administered, randomly by form, to 586 participants in intact classroom settings. There were no between-response direction differences in internal consistency, total score mean, total score variance, or item-to-total correlations. The presence or absence of negatively-worded stems had an effect on Cronbach's alpha, but there was no interaction of this with the response direction variable. In addition, there were no interaction effects associated with sex, age, or handedness orientation on survey statistics. It is concluded that for this survey there was no primacy effect. Reasons why such an effect was not observed when others have detected such an effect are proposed. These relate to differences in types of surveys, the focus of the survey, and the relationship of the topic to the respondent.

While it has not been one of the burning issues in survey design, answering the question of: "Should I use the Likert response alternative pattern Strongly Agree (SA) – Agree (A) – Neutral (N) – Disagree (D) – Strongly Disagree (SD) or the pattern Strongly Disagree (SD) – Disagree (D) – Neutral (N) – Agree (A) – Strongly Agree (SA)?" reflects a practical issue that probably comes up every day. The traditional direction has been SA to SD. A search of the literature and review of the most popular texts on educational measurement provide little guidance for making such a decision. The available research has examined what is referred to as a primacy effect or the tendency for participants to select acceptable options closer to the left side of the response scale. Most of the research on the primacy effect has related to subject recall of presented stimuli and very little has been done to examine primacy effects in the use of Likert-type survey instruments.

In one of the earliest examples of research on this topic, Matthews (1929) concluded that respondents were more likely to select response options to the left rather than the right on a printed survey. Carp (1974) found respondents tended to select responses presented first in an interview situation. The research of others (Johnson, 1981; Powers, Morrow, Goudy, and Keith, 1977) has not generally supported the presence of a primacy effect. Only two recent empirical studies were found (Chan, 1991; Albanese, Prucha, Barnet, & Gjerde, 1997) where self-administered ordered-response surveys were used for the purpose of detecting the primacy effect.

Chan (1991) administered five items from the Personal Distress (PD) Scale, a subscale of the Interpersonal Reactivity Index (Davis, 1980) to the same participants five weeks apart with the first administration using a positive-first response alternative and the second administration using a negative-first response alternative. The alternatives used were variations on "describes me" rather than SD to SA options. Chan found there was a tendency for respondents to have higher scores when the positive-first response set was used, and there were also differences in factor structures between the data sets generated with the two forms of the instrument.

Albanese et al. (1997) used six variations of a student evaluation of instruction form in a medical education setting. The six forms came from crossing the number of response alternatives of five, six, or seven with the response alternative pattern having the "strongly agree" option first or last. They found forms with the most positive statement first (to the left) had more positive ratings and less variance. Of course these statistics are not totally independent when a closed scale is used because as an item mean gets closer to a limit, the variance is constrained.

Neither of these studies, looking at primacy effects, examined possible interaction effects of using all direct or mixed stems or personal characteristics of respondents. Vacha-Haase (1998) pointed out the need for examining score reliability across different studies. The research reported here provides additional empirical evidence for answering the question and examination of possible

---

J. Jackson Barnette is Associate Professor and Associate Dean in the College of Public Health at the University of Iowa. Please address correspondence regarding this article to J. Jackson Barnette, College of Public Health, The University of Iowa, 2811B Steindler Building, Iowa City, IA 52242 or by e-mail to jack-barnette@uiowa.edu.

related variables. More specifically, the research questions are:

1. Are there any differences in internal consistency reliability, total score mean and variance, item means and variances, and item-total correlations between the two response directions?

2. Are any such differences related to the presence or absence of negatively-worded stems or respondent characteristics of sex, handedness orientation, and/or age?

Clearly the researcher would have no control of personal characteristics of respondents and it's not likely that different versions of surveys would be developed such as a form labeled "for use only by left-handed females above 40 years of age." However, had any of these variables been significantly related, there could be great concern and interest in finding out why that may have happened. Simply verifying that these variables are not related to survey internal consistency and score characteristics when comparing differences that may exist between response set directions and use of mixed stems may be useful information to survey researchers. At least it would provide evidence that when used with negatively-worded items such variables were of no concern to the survey researcher. The use of mixed item stems is certainly controlled by the researcher and has been a topic of a great deal of research in the past decades.

Method

Participants

Data were collected from 586 respondents who were asked to complete one of the four forms, assigned randomly, in intact classroom settings. While random selection of participants was not possible, random assignment was accomplished by mixing the survey forms and passing them out as participants entered the room or after they were seated. Respondents were high school students, undergraduate students, graduate students, and inservice teachers in five geographic locations in two states. Table 1 presents the characteristics of the respondents for each survey form and for the total group. Examination of the percentages for the sex, age, and handedness orientation indicates very similar distributions across the four comparison groups.

Table 1
Respondent Characteristics

Response Set Stems	SD to SA All Positive		SD to SA All Positive		SA to SD Mixed		SA to SD Mixed		Total	
	f	%	f	%	f	%	f	%	f	%
Sex										
Male	49	33.6	49	33.6	40	27.6	38	25.9	176	30.1
Female	97	66.4	97	66.4	105	72.4	109	74.1	408	69.9
Unknown			1		1				2	
Handedness Orientation										
Left	19	13.1	10	6.9	13	9.0	12	8.2	54	9.3
Right	126	86.9	134	93.1	132	91.0	135	91.8	527	90.7
Unknown	1		3		1				5	
Age Category										
Less than 20	24	16.4	24	16.4	28	19.2	26	17.7	102	17.4
20 to 29	93	63.7	90	61.6	83	56.8	100	68.0	366	62.6
30 or higher	29	20.0	32	21.9	35	24.0	21	14.3	117	20.0
Unknown			1						1	

Instrumentation and Scoring

Over the years there has been a debate about the use of even or odd numbered categories and the labeling of a midpoint when an odd number of scale points is used. Originally, Likert (1932) used an odd number of categories and this has been the general practice for most of these types of surveys over the years. Generally, the recommendation is made that the critical issue is the nature of the respondents and the likelihood they could be neutral about the topic (NCS, 1996). The topic chosen for this research, attitude toward year-round schooling, could easily elicit neutral responses and research on the primacy effect deals with the proximity of the response category to the left of the response scale. Thus, it seemed not to be a critical concern, and the traditional use of a five-point scale including a neutral middle point was used.

A twenty-item survey on attitude toward year-round schooling was developed and field-tested with 33 participants. The Cronbach alpha for the original form, which used a Likert response alternative pattern of SD-D-N-A-SA, was .85. Another form was developed using the same 20 stems but a response pattern of SA-A-N-D-SD. Ten of the 20 items were randomly selected for negative wording. These ten items were negatively-worded for one of each of the two direction forms by inserting the underlined word "not" resulting in four forms of the instrument, a two-by-two factor pattern. For example one item written in direct form was "Year-round schooling will make it easier for parents to schedule childcare" and written in the negatively-worded form was: "Year-round schooling will not make it easier for parents to schedule childcare." Thus, one factor was response set direction (SA to SD or SD to SA) and the second factor, which was crossed with the response direction factor, was all positively-worded stems and mixed (half positive and half negatively-worded) stems. Additional classification variables, which might be associated with the direction and stem type variables, were respondent sex, handedness orientation, and age. Questions were included at the end of the survey to obtain data on these variables.

Responses were converted to digits 1 to 5 and all scored in the same direction such that higher scores represented more positive agreement with the direct form of the stem. Thus, the negative stem item scores were reflected and the SA to SD responses were reflected to be in the same order as the SD to SA responses.

Data Analysis

Data were analyzed using programs from SAS (1989-1996) including PROC CORR, PROC UNIVARIATE, and PROC GLM. Means were compared using factorial ANOVA, variances were compared using the Brown-Forsythe homogeneity of variance test (Ramsey, 1994), and Cronbach alpha values were compared using equations provided by Feldt, Woodruff, and Salih (1987). The equation for testing interactions involving alpha using a z test was provided by Feldt (personal communication, September 8, 1997) as an extension of the equations found in Feldt et al.

Results

Table 2 provides the results relative to the three primary dependent variables: Cronbach alpha, total mean score (sum of items) response, and standard deviation of total scores. Although standard deviations are reported in the table, actual inferential tests of total score variability (Brown-Forsythe) used variances.

Table 2
Cronbach Alpha, Means, and Standard Deviations
by Response Direction and Stem Configuration

		SD to SA	SA to SD	Total
All positive or direct worded stems	*n*	146	147	293
	α	.8189	.8106	.8154
	M	60.185	60.939	60.563
	SD	9.504	9.506	9.500
Mixed stems, Half positive, Half negative	*n*	146	147	293
	α	.7268	.7063	.7161
	M	62.397	61.776	62.085
	SD	8.344	8.267	8.297
Total	*n*	292	294	586
	α	.7771	.7604	.7690
	M	61.291	61.357	61.324
	SD	8.996	8.903	8.942

There was no statistically significant difference in alpha values when comparing them between the two response alternative directions, $\chi^2(1, n = 586) = 0.3445$, $p = .557$, (.7771 for SD to SA and .7604 for SA to SD). There was a statistically significant difference between the all-positive stem alpha (.8154) compared with the alpha from the mixed-stem instruments (.7161), $\chi^2(1, n = 586) = 12.1282$, $p = <.001$. The difference was statistically significant, therefore it is important to present a measure of practical significance or effect size. Since alpha is a variance-accounted-for statistic, there is a practical difference of .0993 or 9.93% of the total variance. Thus, use of all-positive stems results in about ten percent higher accounting of systematic variance as compared with the use of mixed-stems. While it is clear the alpha values were about ten percent lower when mixed-stems were used compared with having all direct-worded stems, this difference was not related to any type of interaction effect with the response alternative direction variable. This difference is consistent with other research comparing the Cronbach alpha values in the absence and presence of negatively-worded stems (Benson & Hocevar, 1985; Schriesheim & Hill, 1981).

The most important interaction test involving alpha was a test of the interaction of response set direction and absence or presence of negatively-worded items. The alpha values ranged from .7063 to .8189 for the four cells. There was not a statistically significant interaction, z = 0.09, p = .928. There were no statistically significant interactions of response direction with: sex (range of alphas of .7217 to .8049), handedness orientation (range of alphas of .7572 to .8249), or age category (range of alphas of .7355 to .7848).

There was no statistically significant difference between the total score means between the two response directions (61.291 for SD to SA and 61.357 for SA to SD), F (1, 582)= 0.01, p = .929. There was a statistically significant difference between the mean of the all-positive stem group (60.563) and the mixed-stem group (62.085), F(1, 582) = 4.27, p = .039. However, practical significance was very low as indicated by η^2 = .007, and this was not a hypothesis of interest in this research. There was no statistically significant interaction of response direction and absence or presence of negatively-worded items, F (1, 582) = 0.87, p = .351.

There were no statistically significant interactions of response direction with: respondent sex, F (1, 569)= 0.05, p = .819; respondent handedness orientation, F (1, 569)= 0.08, p = .777; or respondent age, F (1, 569)= 1.49, p = .216. Thus, there were no statistically significant mean differences between the two response directions and no statistically significant mean differences in the presence or absence of negatively-worded items or interactions of these two variables with respondent characteristics of sex, handedness orientation, or age.

Based on the results of the Brown-Forsythe test of homogeneity of variance, there was no statistically significant difference between the total score variances between the two response directions, F (1,584)= 0.02, p = .892. Interaction tests for variance were conducted using an omnibus F test of the cell variances. Had that been statistically significant, simple effect tests would have been conducted. There was no statistically significant difference among the four cell variances in the response direction and absence or presence of negatively-worded items configuration, F (3, 582)= 1.91, p = .127. There was no statistically significant difference among the four cell variances in the response direction and respondent sex configuration, F (3, 580)= 0.56, p = .640; there was no statistically significant difference among the four cell variances in the response direction and handedness orientation configuration, F (3, 577)= 0.34, p = .798; and there was no statistically significant difference among the six cell variances in the response direction and age category configuration, F (3, 579)= 0.31, p = .909.

Characteristics of items were examined including item means, standard deviations, item-to-total score correlations. Table 3 presents item means, item standard deviations, and item-total correlations for the two Likert response directions, ranked high to low. There was very high consistency of the item means even when items had negative stems compared with results on those items when they had positive stems. For the entire set of 20 items the Spearman correlation between rank-orders of item means was .979 (p < .001), indicating virtually the same pattern of item means across the two response directions. There was very high consistency of the item standard deviations even when items had negative stems compared with results on those items when they had positive stems. For the entire set of 20 items the Spearman correlation between rank-orders of item standard deviations was .882 (p < .001), indicating virtually the same pattern of item standard deviations across the two response directions. Also, there was very high consistency of the item-to-total score correlations even when items had negative stems compared with results on those items when they had positive stems. For the entire set of 20 items the Spearman correlation between rank-orders of item-to-total correlations was .914 (p < .001), indicating virtually the same pattern of item-to-total correlations across the two response directions.

Conclusions and Discussion

A primacy effect would be manifested in some effect on internal consistency, total score, total score variation, or order of item means, item standard deviation, or item-total correlations. None of these were observed in this experiment. The only statistically significant effect found in this study was that the use of negatively-worded stems had lower, by about .1 (which equates to a ten-percent lower level of systematic measurement variance), values on Cronbach's alpha, but this difference was not related to the response set direction. There were no statistically significant differences in alpha, total score mean, or total score variance between the two Likert response directions and there were no statistically significant interactions with the variables of presence or absence of negatively-worded items, sex of respondent, handedness orientation of respondent, or age of respondent.

Table 3

Means and Ranks of Item Means, Standard Deviations, and Correlations with Total Score for Likert Response Directions

	Item Means				Item Standard Deviations				Item-Total Correlations			
	SD to SA		SA to SD		SD to SA		SA to SD		SD to SA		SA to SD	
Item	*M*	*R*	*M*	*R*	*M*	*R*	*M*	*R*	*M*	*R*	*M*	*R*
1*	2.983	12	3.003	12	1.185	2	1.221	2	.594	5	.499	11
2	2.473	18	2.378	18	1.050	10	1.037	9	.602	4	.604	4
3	3.205	8	3.170	8	1.121	3	1.168	4	.717	1	.735	1
4	3.130	9	3.160	9	1.098	4	1.156	5	.533	11	.601	5
5*	3.418	7	3.507	4	1.089	5	1.089	8	-.041	20	.038	18
6*	4.147	1	4.204	1	0.986	12	0.942	17	.167	16	.050	17
7*	2.908	13	2.915	13	0.920	19	0.929	19	.515	12	.559	9
8*	3.127	10	3.031	10	1.131	7	1.172	3	.043	19	-.027	19
9	3.733	2	3.609	3	0.965	15	0.998	13	.140	17	.214	16
10	3.565	3	3.483	6	1.048	11	1.114	6	.649	2	.705	2
11*	2.692	15	2.755	15	1.222	1	1.229	1	.339	15	.303	15
12*	2.986	11	3.010	11	0.877	20	0.962	15	.583	7	.589	7
13	2.558	17	2.524	17	0.984	13	0.944	16	.461	13	.402	14
14*	3.558	4	3.704	2	1.081	8.5	0.983	14	.072	18	-.088	20
15	2.240	20	2.272	19	0.962	16	1.002	12	.457	14	.478	13
16	2.342	19	2.207	20	1.081	8.5	1.036	10	.624	3	.618	3
17*	2.818	14	2.912	14	0.983	14	1.028	11	.586	6	.568	8
18*	3.421	5.5	3.503	5	0.933	17.5	0.934	18	.576	8	.538	10
19	3.421	5.5	3.439	7	0.933	17.5	0.887	20	.552	10	.497	12
20	2.565	16	2.571	16	1.084	6	1.108	7	.564	9	.596	6
M	3.065		3.068		1.037		1.047		.437		.424	

* indicates item stem was in negative form on mixed versions of survey

Based on these results, there is no evidence that the directionality of Likert response alternatives should be a concern in the design of at least some types of surveys. While this may or may not be an issue for many survey designers, it is a question frequently asked by those learning to design such surveys, and perhaps it is in the back of the minds of many seasoned survey designers. A primacy effect was not observed in this experiment. This indicates that at least sometimes it may not make any difference which direction is used as related to the technical adequacy and stability of the results obtained.

In answer to the original question: "Does it make a difference?", it didn't in this situation. Not only did it not make a difference overall, the variable of whether negatively-worded items were absent or present made no difference. In addition, personal characteristics of sex, handedness orientation, and age were not related to differences in response patterns under either response direction condition. The relative order of item means, the item standard deviations, and item-to-total correlations was highly consistent between the two response directions.

Why were these results not supportive of the findings of Chan and Albanese et al.? There are several possibilities that should be considered. There seem to be two, somewhat related, issues here that may make a difference. First is the nature of the survey itself. The survey used in this research was attitude toward year-round schooling. The items were not ones that would be expected to elicit strong emotional responses one way or another for most respondents. The survey used in the Albanese et al. research was a course evaluation survey, which are notorious for extreme, usually positive, responses. Surveys, which have items that are likely to have responses close to one of the extremes, may be more likely to be prone to a primacy effect. There may be a mental regression effect happening in these cases. There may be a difference between surveys used to collect general opinions as opposed to being used to evaluate something or someone. Related to this may be the nature of what is being evaluated. It may make a difference if the evaluation is of an inanimate object such as a product or project as opposed to a person such as a classroom instructor or supervisor.

Second is the direct personal involvement of the respondent. In Chan's research, respondents were rating themselves on issues that could have been emotionally reactive; they were evaluating themselves. The nature of such self-assessment may be more prone to a primacy effect compared with assessment or attitude toward something or someone other than self. Acquiescence or the provision of socially desirable responses may also be an issue in this and similar situations.

Thus, primacy effect may be determined by situation and assessment strategy rather than by the structure of Likert response alternatives. In a way similar to the assessment of reliability, a survey or test in and of itself is not reliable. Only the scores generated in a specific use of the instrument possess the property of reliability (Thompson, 1994). This may be the case relative to primacy as well. These and other issues remain to be addressed in future research. This study should be replicated using different surveys and respondent types to confirm or refute these findings. However, for most survey applications where highly emotional issues are not addressed, it is likely that the direction of the response sets is not an issue. An issue of greater concern, which this research illustrates, would be the use of mixed item stems or the use of negatively-worded stems.

References

Albanese, M., Prucha, C., Barnet, J. & Gjerde, C. (1997). The direction of positive-rating placement on Likert-type items used to obtain medical students' ratings of instruction. *Academic Medicine, 72*, 627-630.

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary children. *Journal of Educational Measurement, 22*(3), 231-240.

Carp, F. (1974). Position effects on interview responses. *Journal of Gerontology, 29*, 581-587.

Chan, J. (1991). Response-order effects in Likert-type scales. *Educational and Psychological Measurement, 51*(3), 531-540.

Davis, M. (1980). A multidimensional approach to individual differences in empathy. *Catalog of selected documents in psychology, 10*, 85.

Feldt, L., Woodruff, D., & Salih, F. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*(1), 93-103.

Johnson, J. (1981). Effects of the order of presentation of evaluative dimensions for bipolar scales in four societies. *The Journal of Social Psychology, 113*, 21-27.

Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 44-53.

Matthews, C. (1929). The effect of the printed response words on an interest questionnaire. *Journal of Educational Psychology, 30*, 128-134.

National Computer Systems (1996). *Response scales: How many points and what labels* (On-line). Available: http://www.ncs.com/ncscorp/research/96-79.htm.

Powers, E., Morrow, P., Goudy, W., & Keith, P. (1977). Serial order preference in survey research. *Public Opinion Quarterly, 41*, 80-85.

Ramsey, P. (1994). Testing variances in psychological and educational research. *Journal of Educational Statistics, 19*(1), 23-42.

SAS (1989-1996). SAS Institute, Inc.: Cary, NC.

Schriesheim, C., & Hill. K. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement, 41*(4), 1101-1114.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54*, 837-847.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*(1), 6-20.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 1, 83-89

Basic Cross-Validation: Using the "Holdout" Method to Assess the Generalizability of Results

Raquel M. Oxford
University of North Texas

Larry G. Daniel
University of North Florida

Although replication is crucial, it can often involve time-intensive procedures. This is problematic considering that the timeliness of reporting of noteworthy findings is essential if such findings are to make an impact on educational theory and practice. One straightforward procedure for obtaining a more immediate estimate of replicability within the constraints of a single study is the "holdout" or "cross-validation" method: splitting a given sample into relatively equivalent subsamples and comparing results obtained across the subsamples. In this heuristic example, a data set from Holzinger and Swineford (1939) is used to illustrate the method in the multiple regression case. Regression weights obtained with one data subset are used to compare estimated dependent variable scores from the opposite subset.

Replication has been referred to as the hallmark of the social sciences (Carver, 1978). Replication is frequently given lip service, but rarely do researchers take careful precautions to truly replicate important studies and/or to otherwise estimate the degree to which their results are likely to generalize to other populations. When educational researchers use correlational statistical techniques for obtaining scientific results, researchers should be appropriately concerned with the external validity (i.e., the generalizability) of obtained results with respect to a broader population of interest. Because there is always the possibility that results of such analyses will capitalize upon chance, it is desirable that researchers attempt to replicate research findings across various samples selected from a given population (Vockell & Asher, 1995). Hence, replicability has been regarded as the *sine qua non* of research design (Campbell & Stanley, 1963).

Replication efforts may frequently be time consuming and costly. Although the difficulty of conducting replication studies clearly does *not* preclude the necessity of conducting them (certainly there is *no* substitute for replication studies), it is often desirable for researchers to gain an estimate of result generalizability within the limitations of a single study both to determine the degree to which the given sample may be biased and, thereby, to gain an initial idea as to the degree to which the results of the given study are likely to reflect results that would be obtained given other similar samples from the population of interest.

The present article focuses upon one preliminary means for gaining estimates of result generalizability within a single study, namely, the cross-validation "holdout" method. Following a brief treatise on views of science, we overview several methods for cross-validating statistical results, giving special focus to the holdout procedure. We then present results of a multiple regression analysis followed up with a cross-validation holdout analysis using actual research data (Holzinger & Swineford, 1939) to illustrate how the average statistically non-technical researcher could easily employ these analyses routinely in his or her research efforts. Explanation then follows of the various coefficients obtained in the holdout analysis.

Views of Science

Although there are perhaps many possible views one may take toward explaining the nature of science, Kerlinger (1986) proposed a bifurcated system focusing on the "static" and "dynamic" views of science. Within the static view, the emphasis "is on the *present state of knowledge and adding to it* and on the present set of laws, theories, hypotheses, and principles" (p. 7, emphasis in

---

The authors appreciate the helpful comments of five anonymous reviewers on a previous version of this manuscript. Raquel Oxford is the Diversity Scholar Lecturer in the Department of Teacher Education and Administration, College of Education University of North Texas. Larry G. Daniel is Professor and Associate Dean in the College of Education and Human Services at the University of North Florida. Correspondence regarding this articles should be addressed to Raquel Oxford, Department of Teacher Education and Administration, College of Education, University of North Texas, P.O. Box 311337, Denton, TX 76203-1337 or by e-mail to oxford@coefs.coe.unt.edu .

original). Within the dynamic (heuristic) view, science is regarded

> more as an *activity*, what scientists *do*. The present state of knowledge is important, of course. But it is important mainly because it is a base for further scientific theory and research. This has been called the *heuristic view*. The word, "heuristic," meaning serving to discover or reveal, now has the notion of self-discovery connected with it. (p. 8, emphasis in original)

This "*heuristic view*" encompasses much of an ideal that many see as a vital part of professionalism in educational research, namely, being a "scholar educator"--a combination of scholarly researcher and educator of educators and future practitioners. It is not enough to expect self-discovery; rather there should be a nurturing of scholars and practitioners in critical statistical theory and practice. Indeed, one "who knows little or nothing about research and statistics needlessly handicaps his or her performance as a professional" (Kirkpatrick & Aleamoni, 1983, p. 156).

It is from that spirit of scholar educator that the present heuristic exploration was conceived. In the last decade there has been a persistent voice of concern that has sought to gradually affect practices of reporting of results, including concerns with statistical significance, effect size, confidence levels, and evidence of replicability (Ferrell, 1992; Kier, 1997; Lane, 1999; McLean & Kaufman, 1998; Thompson, 1997a, 1997b). Replication is indeed a critical part of the research process. Through the replication of results we aim to generalize, increase our confidence in our results, and address sampling bias, as well as deal with the consideration that statistical significance does not address replicability (Thompson, 1989, 1996, 1997a, 1997b).

Replication: Difficulties and Equitable Alternatives

Although replication is crucial, it can involve time-intensive procedures. Indeed, replication can take years. In fact, as previously noted, replication can be not only time consuming but expensive, and in some cases, it may be impossible or unethical to achieve precision in replication. This is problematic considering that the timeliness of reporting of noteworthy findings is crucial if such findings are to make an impact on educational theory and practice. Publishing requirements at many institutions of higher education demand a significant output of material in a timely fashion in order to receive tenure or promotion. Therefore, time is of the essence for researchers who hope to meet these requirements while simultaneously presenting research findings in which confidence may be placed.

Although no procedure, regardless of how well planned and executed, can serve as a replacement for replication or generalizability of findings, various "sample splitting" (or "invariance") procedures have been proposed as next-best substitutes for replication given the confines of a single study (Ferrell, 1992). The "sample splitting" label is used to describe these procedures considering that the procedures involve various logics for estimating degree of result consistency (or invariance) via reconfigurations of a single sample based on (a) dividing the sample into two or more groups and computing the given analyses separately on data from each group, (b) systematically omitting selected cases and recomputing analyses with the remaining data, or (c) statistically simulating a larger data set using the existing data and then drawing multiple samples from the enlarged data set, with each sample serving as the object of a "repetition" of the given analyses.

There are several "sample splitting" procedures one may employ as equitable estimations of replication given the limitations of a single study (Afifi & Clark, 1984; Fox, 1997). These procedures are frequently referred to as "internal replicability" procedures as they employ data from a single available sample within a given study, as opposed to "external replication" (true replication), which involves complete repetition of the study (including the collection of data from one or more samples independent of the data used in the given study).

One straightforward procedure for obtaining a more immediate estimate of internal replicability within the constraints of a single study is the "holdout" or "cross-validation" method: splitting a given sample into relatively equivalent subsamples and comparing results obtained across the subsamples (Afifi & Clark, 1984; Fox, 1997). Beyond this basic cross-validation procedure, researchers have developed a variety of computationally intensive procedures for developing estimates of generalizability based on varying configurations of a single sample. For example, the "jackknife" method (Tukey, 1958) involves repeated computation of a given statistic with one case or a given subset of cases omitted from the data set each time the statistic is computed. The goal of the jackknife procedure is to "average out" the effects of outlying or atypical cases (Efron & Tibshirani, 1993), thereby offering evidence of the generalizability of the results. A related procedure, bootstrapping, extends the usefulness of the jackknife procedure as it allows for repeated resamplings from a single data set with replacement. Bootstrap replications of a given statistical estimator across a maximal number of fluctuations in the original sample from which the bootstrap data are based (Lunneborg, 1983) serve to create a mock sampling distribution of the statistic of interest. Confidence intervals may be computed for this distribution resulting in estimates as to the likely fluctuations of the statistic of interest (Fox, 1997). All of these procedures (holdout, jackknife, and bootstrap) are valuable tools in the researcher's quest for generalizability; however, as previously noted, because all of the methods are sample specific, they can never replace true replication of results.

Utilizing the Holdout Method

In the present article, we provide an example of the holdout method. Our basic aim is to illustrate the simplicity by which this method may be employed in hopes that educational researchers will begin to use it with greater frequency. The jackknife and bootstrap methods, though extremely useful in educational research, are computationally much more sophisticated, and may be difficult for the statistically non-technical researcher to employ and interpret.

The holdout method is the most fundamental and straightforward procedure for obtaining a more immediate estimate of replicability within the constraints of a single study. The researcher simply randomly splits a given sample into relatively equivalent subsamples and compares results obtained across the subsamples (Afifi & Clark, 1984; Fox, 1997). The researcher should keep in mind that these result comparisons constitute an *estimate* of replicability, and actual external replication is preferred whenever possible (Thompson, 1996).

While the holdout method should be a part of every basic course in statistics and research, a review of the method here is warranted. Our review and subsequent data example utilizes multiple regression; however, the procedure may be utilized with any statistical procedure. The reader should be cautioned, however, that even though the process of splitting the sample is used universally in holdout procedures for any particular statistical method (e.g., descriptive statistics, bivariate correlation, analysis of variance, discriminant analysis), the actual procedures for how to assess result invariance will vary with the procedure employed. Because multiple regression focuses heavily on determination of variable weights and estimates of multiple correlation, or prediction, our invariance analysis utilizes regression coefficients and multiple R. When the holdout method is employed with other statistical procedures, the researcher would focus on specific statistical indices germane to interpretation of that given procedure (e.g., the difference between means in analysis of variance, the magnitude of r^2 in bivariate correlation).

In the multiple regression case, the sample to be employed for the holdout procedure is randomly split into two (or more) roughly equally-sized subsamples (invariance subsamples), or, alternately, invariance subsamples may be based on logical divisions of the data into subsets (e.g, male versus female) regardless of whether the subsample sizes are equivalent. In fact, Thompson (1994) notes that even when random subsamples are drawn, the researcher may wish to make the size of the subsamples disproportionate to increase confidence in results (e.g., a subsample of only 25% of cases that yields consistent results would suggest more confidence in the findings than a subsample consisting of 50% of the results). Regression analyses are run separately for data from each subsample, and regression coefficients (*a* and *b* weights) are determined. These *a* and *b* weights are, respectively, the additive and multiplicative weights used in determining the predicted estimates of the dependent variable ($\hat{Y}$) using the regression predictive equation. For a simple (one predictor variable) regression analysis, the equation is:

$\hat{Y} = a + bX + \text{error}$,
where $\hat{Y}$ = the estimate of the dependent variable (Y),
X = the value of predictor variable, and
error = the difference between the observed and predicted values of the dependent variable ($Y - \hat{Y}$).

In the multiple (two or more predictor variable) regression case, the equation is specified:

$\hat{Y} = a + b_1X_1 + \ldots + \ldots a + b_kX_k + \text{error}$,
where $\hat{Y}$ = the estimate of the dependent variable (Y),
k = the number of predictor variables,
X_1 to X_k = the value of predictor variables 1 through k, and
error = the difference between the observed and predicted values of the dependent variable ($Y - \hat{Y}$).

Once the *a* and *b* weights are known for each subsample, predicted dependent variable scores ($\hat{Y}$) may be artificially calculated for any data subsample using the *a* and *b* weights derived for any other subsample. The question is whether or not one will get a similar result ($\hat{Y}$ value) from the two sub-samples. The researcher hopes to find that the results are sample invariant (i.e., that the results are about the same) whether the predicted dependent variable scores are generated based on a subsample's own regression coefficients or the regression coefficients obtained using a different subsample's data. Invariance of results indicates that the sample used to develop the regression weights is unimportant so long as the samples are drawn from the same population. The ideal situations for use of the holdout method are with a relatively large sample and when there is an appreciable R^2, at least 0.10.

An excellent example of how meaningful the holdout technique can be for the educational practitioner and researcher was presented in *Experimental Research in Counseling* (Kirkpatrick & Aleamoni, 1983, p. 116):

> One can **cross-validate** [emphasis in the original] the multiple regression equation by using it to predict the values on the dependent variable for a second sample and then finding a new multiple R, correlation between the real Y values and the predicted values, Y'.
>
> [Suppose, for example that a]. . . high school counselor . . . wanted to predict freshman college GPA for high school seniors using the seniors' high school GPA, entrance exam scores, and other data available on the students. . . . The hypothetical counselor used these same data for last year's high school graduates—and their known freshman GPA's from college—to construct a multiple regression equation for predicting freshman college GPA's for this year's high school seniors. . . .
>
> The high school counselor could cross-validate by dividing the group of last year's seniors into halves. Then, the multiple regression is done using the values from one half and the equation used to "predict" the (known) freshman college GPA's for the other half. Through the [cross-validation] process the counselor could gauge the amount of error involved in applying the equation to other samples.

Methodology

The data set employed herein to illustrate basic cross-validation with the holdout method in the multiple regression case was taken from a classic study by Holzinger and Swineford (1939). The data were collected from the administration of 24 tests of ability to a group of 301 middle school students. The 24 tests can be divided into five major sets, each with its own ability theme—spatial, verbal, speed, memory, and math. This data set has been used with a good bit of frequency in educational research situations showing applications of various statistical procedures (e.g., Crowley & Thompson, 1991; Thompson & Daniel, 1991). Because the present analyses were conducted for illustrative purposes only, no substantive interpretation of the data is presented or implied.

All regression analyses were run using the regression procedure in SPSS 9.01 specifying four predictor variables. The variable names and labels are shown in Table 1, and the tests selected for this demonstration are the tests associated with verbal performance. The dependent variable was T6, the Paragraph Comprehension Test. The first regression analysis was run using the full sample (*N* = 301). Next, the analysis was run again with roughly half (*n* = 151) of the cases in a systematic random subsample comprised of the odd numbered cases. Then the other half (*n* = 150) representing the invariance group with the even numbered cases was analyzed. Predicted dependent variable ($\hat{Y}$) scores were saved for each half and renamed $\hat{Y}_{ODD}$ ("yhatodd") and $\hat{Y}_{EVEN}$ ("yhateven"). In actual practice, one of these sample halves would have been specified as the "research" subsample, and the other would have been deemed the "holdout" subsample.

Table 1
Codes for Data From Holzinger and Swineford (1939)

T5	GENERAL INFORMATION VERBAL TEST
T6	PARAGRAPH COMPREHENSION TEST
T7	SENTENCE COMPLETION TEST
T8	WORD CLASSIFICATION-WHICH WORD NOT BELONG IN SET
T9	WORD MEANING TEST

Regression weights obtained with one data subset were used to compute estimated variable scores ($\hat{Y}$) using cases from the opposite subset. These "cross-weighted" $\hat{Y}$ estimates can be computed with relative ease even by the statistically nontechnical researcher using the printout showing the regression *a* and *b* coefficients for the regression analysis using each subsample's data. The regression equations can be quickly composed, and cross-weighted $\hat{Y}$ values may be calculated using a spreadsheet or a computation command available in most statistical software packages (e.g. the Transform/Compute procedure available in SPSS).

Results

The initial regression results and the coefficients yielded by the analysis using the entire sample of 301 students are shown in Tables 2 and 3, respectively. Regression results and resultant coefficients using data from the first subsample (odds) are presented, respectively, in Tables 4 and 5; regression results and coefficients for the second subsample (evens) are shown in Tables 6 and 7, respectively. Using the full sample, results were statistically significant ($p < .001$) and there was a large statistical effect ($R^2 = .612$). When the regression analysis was run separately for each of the individual invariance groups, results were statistically significant ($p < .001$) in each case. Similar to the full sample results, these separate analyses yielded R^2's of .645 and .576, respectively, for the odd- and even-numbered cases.

Table 2
Initial Regression Analysis for Verbal Tests (T5, T7, T8, T9) Predicting Paragraph Comprehension Test (T6)

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	2238.452	4	559.613	116.611	<.001
Residual	1420.498	296	4.799		
Total	3658.950	300			

Table 3
Coefficients for Initial Regression Analysis for Verbal Tests (T5, T7, T8, T9) Predicting Paragraph Comprehension Test (T6)

	Unstandardized Coefficients		Standardized Coefficients		
Model	B	Std. Error	Beta	t	Sig.
(Constant)	6.301E-02	.629		.100	.920
T5	3.074E-02	.017	.109	1.807	.072
T7	.262	.041	.387	6.367	.000
T8	4.713E-02	.032	.077	1.488	.138
T9	.137	.027	.301	5.104	.000

Table 4
Regression Analysis of Odd-Numbered Cases

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1272.341	4	318.085	66.193	<.001
Residual	701.593	146	4.805		
Total	1973.934	150			

Table 5
Coefficients for Regression Analysis of Odd-Numbered Cases

	Unstandardized Coefficients		Standardized Coefficients		
Model	B	Std. Error	Beta	t	Sig.
(Constant)	-.349	.888		-.393	.695
T5	4.068E-02	.023	.137	1.739	.084
T7	.246	.056	.368	4.389	.000
T8	5.695E-02	.043	.090	1.321	.189
T9	.140	.037	.306	3.746	.000

Table 6
Regression Analysis of Even-Numbered Cases

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	969.956	4	242.489	49.253	<.001
Residual	713.884	145	4.923		
Total	1683.840	149			

Table 7
Coefficients for Regression Analysis of Even-Numbered Cases

	Unstandardized Coefficients		Standardized Coefficients		
Model	B	Std. Error	Beta	t	Sig.
(Constant)	.501	.904		.554	.580
T5	2.103E-02	.026	.078	.806	.421
T7	.281	.063	.408	4.437	.000
T8	3.300E-02	.048	.055	.682	.497
T9	.136	.040	.301	3.401	.001

Simply eyeballing these initial results would suggest that the two data subsamples were rather consistent in their degree of predictive accuracy. However, the cross-validation procedure requires further investigation of the findings using various "invariance coefficients." Invariance coefficients are simply statistical estimates that are derived when information taken from one data subsample is used to compute scores for the other subsample. Two types of invariance coefficients were computed for the data in hand. The first set of invariance coefficients were Pearson correlation coefficients (r's) between $\hat{Y}$ values computed for each subsample using its own regression coefficients and those computed using the other subsample's coefficients. These values are referred to herein as "coefficient stability invariance coefficients." The second set of invariance coefficients, herein referred to as "cross-weighted multiple R values" were Pearson correlations between the dependent variable values and cross-weighted $\hat{Y}$ values for a given subsample.

Although there are no firmly established criteria for interpreting the strength of invariance coefficients, the coefficients obtained for the present sample are ideal. The coefficient stability invariance coefficient calculated between the odd-numbered sample's $\hat{Y}$ values using actual and cross-determined coefficients (r between $\hat{Y}_{OBSERVED}$ and $\hat{Y}_{EVEN}$) was .997. Similarly, the coefficient stability coefficients for the even-numbered sample's $\hat{Y}$ values using actual and cross-determined coefficients (r between $\hat{Y}_{OBSERVED}$ and $\hat{Y}_{ODD}$) was .998. These coefficients indicate that the obtained regression coefficients are nearly 100% sample invariant! These results are not necessarily typical, but in the present case, assuming the sample was carefully selected, the results offer some "evidence that the original results are somewhat generalizable to the population of interest" (Ferrell, 1992, p. 16).

The cross-weighted multiple *R* invariance coefficients were equally promising. As previously shown in Tables 4 and 6, the regressions run, respectively, with the odd- and even-numbered cases yielded *R* values of .803 and .759. The same values computed using the cross-weighted $\hat{Y}$ values (i.e., $r_{\hat{Y}Y}$) were .801 and .757, respectively. Hence, there is virtually no difference in the estimates of predictive accuracy regardless of the particular data subsample employed for deriving the coefficient of multiple correlation.

Discussion

As the foregoing results indicate, basic cross validation using the holdout method is a very simple procedure to utilize. In light of the time constraints of replication yet the necessity of some indication of such, this method can be very useful to researchers, scholars, and practitioners who wish to develop initial estimates of the replicability of their findings. The usefulness and computational simplicity of invariance coefficients based on the holdout method prompt our recommendation that educational researchers more frequently utilize these coefficients in reporting their findings. Obviously, as noted previously, use of the holdout procedure, or any other method of cross-validation, cannot replace replication, but as an initial method for estimating result stability as well as identifying possible bias within a given sample, holdout analysis is clearly promising.

It should be noted that the holdout method is not without its limitations. If the sample employed in the cross-validation analysis includes biased or skewed data and/or if the sample does not represent the population, an overestimation or underestimation of the statistic of interest can take place. Nevertheless, a biased estimate of result invariance, is better than no estimate at all (Daniel, 1992), and although the desire to conduct a holdout analysis does not preclude the necessity for the researcher to employ various methods for screening data prior to analysis (American Psychological Association Board of Scientific Affairs, 1999), the holdout method is extremely useful once some reasonable determination of the "goodness" of the sample is established. Another potential problem with the holdout method occurs in cases in which a small sample is employed. Obviously, splitting an already small sample can result in result shrinkage (Daniel, 1992; Reinhardt, 1992; Schmitt, 1989); hence, the holdout procedure should be avoided when sample size is extremely small. The key is that the holdout procedure is an easy way to assess generalizability of results when the foregoing concerns are addressed.

References

American Psychological Association Board of Scientific Affairs. (1999). Statistical methods in psychological journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Afifi, A. A., & Clark, V. (1984). *Computer-aided multivariate analysis*. Belmont, CA: Lifetime Learning Publications.

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48*, 378-399.

Crowley, S. L., & Thompson, B. (1991, April). *Evaluating the replicability of multivariate assessment and evaluation results: A review of various applications of the cross-validation logic*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago. (ERIC Document Reproduction Service No. ED 333 017)

Daniel, L. G. (1992, April). *Bootstrap methods in the principal components case*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 346 135)

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.

Ferrell, C. M. (1992, January). *Statistical significance, sample splitting, and generalizability of results*. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston. (ERIC Document Reproduction Service No. ED 343 935)

Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution* (No. 48). Chicago: University of Chicago.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). Ft. Worth, TX: Holt, Rinehart and Winston.

Kier, F. J. (1997, January). *Ways to explore the replicability of multivariate results (since statistical significance testing does not)*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 407 423)

Kirkpatrick, J. S., & Aleamoni, L. M. (1983). *Experimental research in counseling*. Springfield: Charles C. Thomas Books.

Lane, G. R. (1999, November.). *Show me the magnitude! The consequences of overemphasis on null hypothesis significance testing*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL.

Lunneborg, C. E. (1983, August). *Efron's bootstrap with some applications in psychology*. Paper presented at the annual meeting of the American Psychological Association, Anaheim, CA. (ERIC Document Reproduction Service No. ED 241 554)

McLean, J. E., & Kaufman, A. S. (Eds.). (1998). Statistical Significance Testing [Special Issue]. *RESEARCH IN THE SCHOOLS*, 5 (2).

Reinhardt, B. M. (1992, April). *Estimating result replicability using double cross-validation and bootstrap methods*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 346 164)

Schmitt, D. R. (1989, November.). *Three methods of cross-validation: Is there a difference in conclusions?* Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock. (ERIC Document Reproduction Service No. ED 314 440)

Thompson, B. (1989). Statistical significance, result importance, and result generalizabilty: Three noteworthy but somewhat different issues. *Measurement and Evaluation in Counseling and Development*, *22*, 2-5.

Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality*, *62*, 157-176.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, *25*(2), 26-30.

Thompson, B. (1997a). *If statistical significance tests are broken/misused, what practices should supplement or replace them?* Paper presented at the annual meeting of the American Psychological Association, Chicago. (ERIC Document Reproduction Service No. ED 413 342)

Thompson, B. (1997b). Rejoinder: Editorial policies regarding statistical significance tests: Further comments. *Educational Researcher*, *26*(5), 29-32.

Thompson, B., & Daniel, L. G. (1991, April). *Canonical correlation analyses that incorporate measurement and sampling error considerations*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 334 248)

Tukey, J. W. (1958). Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics*, *29*, 614.

Vockell, E. L., & Asher, J. W. (1995). *Educational research* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

JOURNAL SUBSCRIPTION FORM

This form can be used to subscribe to *RESEARCH IN THE SCHOOLS* without becoming a member of the Mid-South Educational Research Association. It can be used by individuals and institutions.

◆

Please enter a subscription to *RESEARCH IN THE SCHOOLS* for:

Name: ______________________________

Institution: ______________________________

Address: ______________________________

____

			COST
Individual Subscription ($25 per year)	Number of years	_____	_______
Institutional Subscription ($30 per year)	Number of years	_____	_______
Foreign Surcharge ($25 per year, applies to both individual and institutional subscriptions)	Number of years	_____	_______
Back issues for Volumes 1 - 7 ($30 per Volume)	Number of Volumes	_____	_______
TOTAL COST:			_______

MAKE CHECKS PAYABLE TO MSERA
SEND FORM AND CHECK TO:

Dr. James E. McLean, Co-Editor
RESEARCH IN THE SCHOOLS
Warf-Pickle Hall, Room 418
Eastern Tennessee State University
Box 70685
Johnson City, TN 37614-1709

MID-SOUTH EDUCATIONAL RESEARCH ASSOCIATION

The Mid-South Educational Research Association (MSERA) was founded in order to encourage quality educational research in the Mid-South and to promote the application of quality educational research in schools. Members of MSERA share interests in educational research, development, and evaluation. While most members are from institutions of higher education, many others represent state departments of education, public and private agencies, and public school systems. Graduate students comprise a significant portion of the membership. A majority of MSERA members are from the six states represented by the organization, but others are from many other states and several foreign countries. The MSERA is the largest regional educational research organization in the country.

The organization provides several services for its members. The annual meeting, held every November, offers many formal and informal opportunities for professional development through special training courses, sharing of research findings, and programmatic interests with colleagues. Members receive a subscription to *RESEARCH IN THE SCHOOLS* and the *Mid-South Educational Researcher*. The MSERA also provides recognition and cash rewards for its outstanding paper, an outstanding dissertation, and professional service.

✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂✂

MSERA Membership/Renewal Form

(Please print or type)

Name ____

Organization ____

Address ____

____

Telephone Work: ____________________

Home: ____________________

Fax: ____________________

e-mail: ______________________________

Amount Enclosed:	MSERA 2001 Membership ($25 professional, $15 student)	$______
	MSER Foundation Contribution	$______
	TOTAL	$______

Make check out to MSERA and mail to:

Dr. Ernest Rakow
University of Memphis
Ball Education Bldg., Room 100
Memphis, TN 38152-6010

RESEARCH IN THE SCHOOLS
Mid-South Educational Research Association
Warf-Pickle Hall, Room 418
Eastern Tennessee State University
Johnson City, TN 37614-1709

BULK RATE
U.S. POSTAGE
PAID
PERMIT NO. 3029
BIRMINGHAM, AL

MSERA

RESEARCH IN THE SCHOOLS

A nationally refereed journal sponsored by the
Mid-South Educational Research Association
and East Tennessee State University.

Volume 8, Number 2 **Fall 2001**

The Relationship Between Eighth-Grade Reading Scores and Achievement on the Georgia High School Graduation Test 1
Deborah L. Demps and Anthony J. Onwuegbuzie

From Texas to Florida: Email Peer Coaching Dialogues Among Preservice Teachers 11
Susanne I. Lapp

East Baton Rouge, School Desegregation, and White Flight 21
Stephen J. Caldas and Carl L. Bankston III

Science Achievement, Class Size, and Demographics: The Debate Continues 33
Marie Miller-Whitehead

Characteristics of Effective Teachers: Perceptions of Preservice Teachers 45
Ann E. Witcher, Anthony J. Onwuegbuzie and Lynn C. Minor

Teachers' Beliefs About Mathematics Reform: Instructional Implications for Students With Learning Disabilities 59
Kathleen M. T. Collins and Michael M. Gerber

Academic Success as a Function of the Gender, Class, Age, Study Habits, and Employment of College Students 71
William J. Lammers, Anthony J. Onwuegbuzie, and John R. Slate

The Utility of Statistical Significance Testing in Psychological and Educational Research: A Review of Recent Literature and Proposed Alternatives 83
Jeremy R. Sullivan

James E. McLean and Alan S. Kaufman, Editors

RESEARCH IN THE SCHOOLS

Information for Authors

Statement of Purpose

RESEARCH IN THE SCHOOLS (ISSN 1085-5300) publishes original contributions in the following areas: 1) *Research in Practice*--empirical studies focusing on the results of applied educational research including cross-cultural studies, 2) *Topical Articles*--scholarly reviews of research, perspectives on the use of research findings, theoretical articles, and related articles, 3) *Methods and Techniques*--descriptions of technology applications in the classroom, descriptions of innovative teaching strategies in research/measurement/statistics, evaluations of teaching methods, and similar articles of interest to instructors of research-oriented courses, 4) *Assessment*--empirical studies of norm-referenced, criterion-referenced, and informal tests in the areas of cognitive ability, academic achievement, personality, vocational interests, neuropsychological functioning, and the like, and 5) *Other* topics of interest to educational researchers. *RESEARCH IN THE SCHOOLS* is devoted to research conducted in *any* educational setting from a conventional elementary school or high school to a training program conducted within an industry. Likewise, there are no age restrictions on the sample, since the educational settings may include preschools, continuing education classes for adults, or adaptive skills courses in nursing homes. Studies conducted in settings such as clinics, hospitals, or prisons are ordinarily inappropriate for *RESEARCH IN THE SCHOOLS* unless they involve an educational program within such a setting. One goal of *RESEARCH IN THE SCHOOLS* is to provide a training ground for graduate students to learn effective reviewing techniques. Consequently, the journal utilizes a Graduate Student Editorial Board composed mostly of students in educational psychology and educational research. Members of this Editorial Board, each sponsored by a professor, provide supplementary reviews for a selection of submitted articles, and receive both direct and indirect feedback of the quality of these reviews.

Preparing Manuscripts

Authors should prepare manuscripts in accordance with the stylistic rules and guidelines delineated in the *Publications Manual of the American Psychological Association* (4th ed., 1994), which is available from: Order Department, American Psychological Association, PO Box 2710, Hyattsville, MD 20784. Number the pages consecutively. All manuscripts will be subject to editing for sexist language.

Author Identification

Authors should put the complete title of the article on the first text page, but they should exclude their names. Subsequent pages should include only a running head. They should prepare a separate sheet with the complete title of the article and their names and affiliations; this procedure will ensure anonymity in the review process. Authors should supply addresses and phone numbers, and electronic mail addresses and fax numbers (if available), for potential use by the editorial staff and, later, by the production staff. Unless otherwise stated, the first-named author will be sent correspondence, galley proofs, copyright forms, and so forth.

Submission of Manuscripts

Submit manuscripts in triplicate to **James E. McLean, Co-Editor, RESEARCH IN THE SCHOOLS, Warf-Pickle Hall, Room 418, Eastern Tennessee State University, Box 70685, Johnson City, TN 37614-1709. Please direct questions to jmclean@etsu.edu**. All copies should be clear and readable; dot matrix is acceptable only if it meets these qualities of legibility. Length of the manuscripts, including references and tables, should ordinarily range from about 10 to 40 typed, double-spaced, 8-1/2 X 11-inch pages, using 11-12 point type. Abstracts are limited to 125 words. Brief reports of research are not encouraged. Authors are encouraged to keep a hard copy of the manuscript to guard against loss. It is assumed that all manuscripts submitted for publication are original material and have not been simultaneously submitted for publication elsewhere. When manuscripts are accepted for publication, authors are encouraged to submit the final version on a computer disk along with the hard copy.

Copyright and Permissions

Authors are granted permission to reproduce their own articles for personal use. Others must request permission to reproduce tables, figures, or more than 500 words of text from the editors. Copyright © 2001 by the Mid-South Educational Research Association.

RESEARCH IN THE SCHOOLS

Volume 8, Number 2 **Fall 2001**

EDITORS
James E. McLean, *East Tennessee State University*
Alan S. Kaufman, *Yale University School of Medicine*

ASSOCIATE EDITORS
Charles M. Achilles, *Eastern Michigan University*
James C. Kaufman, *Educational Testing Service*

PRODUCTION EDITOR
Margaret L. Rice, *The University of Alabama*

COPY EDITOR
Michele G. Jarrell, *Shelton State Community College*

EDITORIAL BOARD
Jan E. Alm, *Uppsala University (Sweden)*
Margery Arnold, *Trinity University*
Mark Baron, *University of South Dakota*
Alan W. Brue, *Carrollton City (Georgia) Schools*
Robin Cook, *Wichita State University*
Larry G. Daniel, *University of North Florida*
James M. Ernest, *State University of New York at Buffalo*
Daniel Fasko, Jr., *Ohio University*
Ann Gallagher, *Educational Testing Service*
Glennelle Halpin, *Auburn University*
Harrison Kane, *Western Carolina University*
JinGyu Kim, *National Institute of Professional Administration (Korea)*
Jwa K. Kim, Middle *Tennessee State University*
Frank R. Lawrence, *Pennsylvania State University*
Robert E. Lockwood, *Department of Postsecondary Education (Alabama)*
Robert Marsh, *Chattanooga State Tech. Community College*
Jerry Mathews, *Auburn University*
Charles McLafferty, Jr., *University of Alabama at Birmingham*
Peter C. Melchers, *University of Cologne (Germany)*
Claire Meljac, *Hospital Sainte-Anne (France)*
Soo-Back Moon, *Catholic University of Hyosung (Korea)*
David T. Morse, *Mississippi State University*
Jack A. Naglieri, *George Mason University*
Anthony J. Onwuegbuzie, *Howard University*
Cecil R. Reynolds, *Texas A & M University*
Janet C. Richards, *The University of Southern Mississippi*
Michael D. Richardson, *Georgia Southern University*
John Kyle Roberts, *Baylor College of Medicine*
Dan Robinson, *University of Texas at Austin*
Christopher Skinner, *University of Tennessee*
John R. Slate, *University of Texas at El Paso*
Paula Stanley, *Radford University*
Bruce Thompson, *Texas A & M University*

GRADUATE STUDENT EDITORIAL BOARD
Keri L. Altig, *University of Nevada, Las Vegas*
Adam S. Bristol, *Yale University*
James M. Graham, *Texas A & M University*
Debbie Hahs, *The University of Alabama*
Keith Hyatt, *University of Nevada, Las Vegas*
Kevin M. Kieffer, *Texas A & M University*
Ginny Lane, *University of North Texas*
Lynn C. Minor, *Valdosta StateUniversity*
William T. Pepper, *Mississippi State University*
Catherine A. Stacy, *University of Texas at Austin*

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 2, 1-9

The Relationship Between Eighth-Grade Reading Scores and Achievement on the Georgia High School Graduation Test

Deborah L. Demps
Valdosta State University

Anthony J. Onwuegbuzie
Howard University

The purpose of the study was to determine if eighth-grade reading scores on the Iowa Tests of Basic Skills (ITBS) predict success on the five subtests (i.e., Writing, Language Arts, Math, Social Studies, and Science) of the Georgia High School Graduation Test (GHSGT). Findings revealed that reading scores were statistically significantly related to all five subtests, with correlations ranging from .69 to .86. Although these correlations were similar between Caucasian-American and minority students, some gender differences emerged. Specifically, the correlations between reading scores and scores on the Writing, Mathematics, and Social Studies subtests were statistically significantly stronger for females than for males—with effect sizes (i.e., differences in Fisher's z values) of .43, .50, and .43, respectively. Finally, a series of independent samples t-tests revealed that students who passed a subtest of the GHSGT (i.e., obtained a score of 500 points or more) had attained statistically significantly higher eighth-grade reading scores than did their failing counterparts. All effect sizes were extremely large. The educational implications are discussed.

The 1991 Georgia General Assembly established that in addition to earning 24 Carnegie units, high school students in Georgia must pass all portions of the Georgia High School Graduation Test (GHSGT) in order to earn a regular education diploma (Georgia Department of Education, 1991). Based on learning objectives specified in the Georgia Quality Core Curriculum for Grades 9 through 12, the GHSGT comprises subtests in the areas of Language Arts, Writing, Mathematics, Social Studies, and Science. Students begin taking the GHSGT during the spring of their junior year and have five opportunities before the end of their 12th-grade year to retake portions that they may have failed.

Many researchers have investigated factors which influence academic achievement. For example, Weller and Weller (1997) concluded that high school students' academic abilities can be directly linked to their reading achievement. In addition, Simner and Barns (1991) found that students who displayed weaknesses in reading comprehension in early grades usually experienced academic problems at the secondary level. Although not yet empirically tested, it is likely that lack of sufficient reading skills may, in part, be responsible for poor performance on the GHSGT. Therefore, the purpose of the study was to determine if eighth-grade reading scores on the Iowa Tests of Basic Skills (ITBS) predict success on the five subtests (i.e., Writing, Language Arts, Math, Social Studies, and Science) of the Georgia High School Graduation Test (GHSGT). Although researchers have investigated the use of standardized tests to predict academic success at the postsecondary level, limited research exists on the use of standardized tests to predict academic performance at the high school level. Moreover, an extensive review of the literature revealed no study in which correlates of scores on the GHSGT were investigated. Bearing in mind that every student seeking a high school diploma from a Georgia school must pass all five subtests of the GHSGT (Georgia Department of Education, 2000), there is a dire need for information that will help to identify students who are at risk of failing the GHSGT.

Research Question

The following research question was addressed in this investigation: What is the relationship between reading scores on the Iowa Tests of Basic Skills (ITBS) and scores on all five subtests of the Georgia High School Graduation Tests (GHSGT)?

Hypotheses

The following five hypotheses were tested:

Deborah L. Demps is the Curriculum/Title I Director at Brooks County (Georgia) School System. She is also a doctoral candidate at Valdosta State University. Anthony J. Onwuegbuzie is an associate professor in the Department of Human Development and Psychoeducational Studies at Howard University. Please address correspondence to Anthony Onwuegbuzie, Department of Educational Leadership, College of Education, Valdosta State University, Valdosta, Georgia, 31698 or by e-mail to TONWUEGB@VALDOSTA.EDU.

Hypothesis 1. There is a positive relationship between eighth-grade reading scores on the Iowa Tests of Basic Skills and scores on the Georgia High School Graduation Writing Test.

Hypothesis 2. There is a positive relationship between eighth-grade reading scores on the Iowa Tests of Basic Skills and scores on the Georgia High School Graduation Language Arts Test.

Hypothesis 3. There is a positive relationship between eighth-grade reading scores on the Iowa Tests of Basic Skills and scores on the Georgia High School Graduation Math Test.

Hypothesis 4. There is a positive relationship between eighth-grade reading scores on the Iowa Tests of Basic Skills and scores on the Georgia High School Graduation Social Studies Test.

Hypothesis 5. There is a positive relationship between eighth-grade reading scores on the Iowa Tests of Basic Skills and scores on the Georgia High School Graduation Science Test.

Review of Related Literature

The recent emphasis on standardized tests has sparked many debates in testing policies and procedures on the national, state, and local levels. For instance, Waltman (1997) suggested that it might be beneficial to state legislators if statewide assessments were linked to a set of national standards such as the National Assessment of Educational Progress (NAEP), provided that this can be accomplished without introducing unacceptably large errors. To explore this possibility, Waltman used an equipercentile procedure (a procedure used to establish a link between scores on two or more assessments) to determine the extent to which performance on the Iowa Tests of Basic Skills (ITBS) and the NAEP mathematics scale score represented the same achievement levels. The equipercentile link produced percentages on the ITBS scale that were similar to those reported by NAEP for the following four subgroups that were classified by type of community: advantaged urban (n = 561), disadvantaged urban (n = 429), extremely rural (n = 1,994), and other nonextreme (n = 2,935). For students taking both the NAEP and ITBS mathematics tests, low to moderate percents of agreement in achievement-level classifications (i.e., below basic, at or above basic, at or above proficient, at or above advanced) were found. These results led Waltman to conclude that the nature of the relationship between the NAEP and ITBS scales was "vague…[and] cast doubt on the appropriateness of maing inferences with respect to group performance" (p. 118).

Unless there is a federal intervention to impose a national assessment, it appears likely that states will continue to use their own forms of assessment. In fact, many states have enacted legislation requiring students to demonstrate basic skills by achieving a passing score on a minimum competency test. As a result, emphasis on standardized tests has increased at the state level. For example, in an effort to raise standardized test scores, some school systems have begun using programs designed to align their state curricula with the skills measured by the standardized tests used in the district (i.e., curriculum alignment programs) (Gandal, 1995). On the contrary, Brent and DiObilda (1993) indicated that the direct-instruction program was just as effective as the curriculum alignment program in improving student achievement.

Selected Debates Over Standardized Testing

Although the use of standardized tests in education has been well documented, researchers disagree on their use as the sole assessment of academic achievement. Opponents of standardized testing have proposed that performance-based assessments would be a more appropriate measurement of student achievement. Strong and Sexton (1996) addressed this contention by comparing scores on Kentucky's performance-based assessment, Kentucky Instructional Results Informational System (KIRIS), and reading scores on the American College Test (ACT). Findings from this study revealed substantial disagreement between KIRIS and ACT reading results. Approximately 29% of the students who scored at the "Novice" level on the KIRIS achieved average and above average scores on the ACT. Similarly, 64% of students who scored at the "Apprentice" level on the KIRIS achieved average and above scores on the ACT. As a result, Strong and Sexton concluded that the use of the KIRIS for state accountability was questionable.

Charlesworth, Fleege, and Weitman (1994) contended that the heavy reliance on testing has created schools in which "test-driven" instruction is a common practice. Indeed, Charlesworth et al. (1994) noted that (a) the United States spends $500,000,000 annually on testing, which accounts for nearly 100 million tests administered each year; (b) over 20 million school days are delegated to testing each year; and (c) in some cases, the same students are tested between 7 and 12 occasions in any given year. According to these researchers, the pressure from administrators and the community has resulted in teachers spending instructional time preparing students for tests. This widespread tendency of teachers to "teach the test," however, produced a negative educational impact that was more severe for minority students, particularly those who reside in the inner city (National Commission on Testing and Public Policy [NCTPP], 1990). Lomax, West, Harmon, Viator, and Maddaus

(1995) and Herman, Abedi, and Golan (1994) concluded that classes in which "teaching to the test" occurred were more detrimental to minority students because more emphasis was placed on drill and practice skills than on higher-order thinking skills, conceptual knowledge, and procedural knowledge. The authors concluded that these skills are essential to success at the secondary and post-secondary levels, and if they are not taught, the success of minority students at these levels may decrease.

Schmeiser and Ferguson (1979) addressed the contention that some tests are biased in ways that are unfair to minorities. To investigate the effect of test content on African-American and Caucasian-American students' performance levels, the researchers used specially-developed tests which reflected both African-American and Caucasian-American cultures. All tests were developed to measure the same cognitive skills and were representative of skills taught nationally in high schools and colleges. Specifically, the 18-item test developed for the African-American sample was based on an African-American Woman's reflections on her Civil War experiences. On the other hand, the 17-item test developed for the Caucasian-American sample was based on content associated with the Caucasian-American culture, including topics such as English literature, the U.S. fight for independence, and motorcycling. Schmeiser and Ferguson found that as a group, Caucasian-American students scored significantly higher than did African-American students on both versions of the tests. Surprisingly, the cultural content assessed on the tests did not differentially affect either group's test performance.

Some researchers also have questioned the validity of the use of standardized tests for placement and selection of students for advanced and remedial classes. According to Espin and Deno (1993), standardized tests are inadequate for making decisions about student placement. Similarly, Levande (1993) contended that standardized test scores are not valid measures of academic ability because they only report mastery of certain concepts. Moreover, Levande asserted that because factors that adversely affect test results (e.g., poor health, problems at home) are often ignored by school authorities, and because test results are computed statistically as single scores, standardized test scores distort reality and are overemphasized.

High School Competency Tests and Achievement

The use of state-mandated graduation examinations also has been a concern of researchers (Jacobson, 1998). Critics of state graduation tests disagree with the view that high stakes tests improve student achievement. Indeed, authors of a report from the National Center for Fair and Open Testing, or FairTest, concluded that students in states with high school exit examinations are less likely to meet proficiency levels on the National Assessment of Educational Progress (NAEP) math and reading tests (Jacobson, 1998). On the other hand, researchers such as Lessitz (1997) favor the use of standardized testing to certify students for graduation. According to Lessitz, graduation assessments ensure that students have the basic skills for employment.

Results of a study by Griffin and Heidorn (1996) did not support the hypothesis that minimum competency tests have an adverse effect on at-risk, disadvantaged students. These researchers concluded that while academically-disadvantaged students are more likely to leave school, it did not appear that performance on minimum competency tests provided any additional impetus for these students to drop out of high school. A risk ratio was used on 76,664 students enrolled in 75 high schools in the state of Florida to compute the students' probability of dropping out of high school. These probabilities revealed that as academic performance increased, the predicted probability of dropping out of school decreased. However, at lower levels of grade point average (GPA), no statistically significant differences were found between students who passed and failed the minimum competency test. Surprisingly, however, a statistically significant higher risk of leaving school was found for students with higher GPAs who did not pass the minimum competency test than for those who passed the test.

Norton and Park (1996) used data from 10th- and 11th-grade students who participated in the spring 1996 administration of the Louisiana Graduation Exit Exam to investigate the relationship between students' test preparation and their academic performance. The researchers used a survey question for each subject area of the test (i.e., English Language Arts, Mathematics, Written Composition, Science, and Social Studies) to measure students' test preparation. Academic ability was assessed through students' pass status and their scale scores. The researchers found that students' test preparation was significantly associated with their passing status on all five subject areas. The magnitudes of the relationships were different across all subject areas; however, the strongest relationship was found in the area of Mathematics.

Standardized Tests as Predictors of Academic Success

Some researchers have debated the usefulness of standardized tests as a predictor of students' academic success in college. For example, Baron and Norman (1992) and Sacks (1997) concluded that the Scholastic Achievement Test (SAT) and the Graduate Record Exam (GRE) made relatively small contributions as a predictor of success in college. In addition, Baron and Norman found that the SAT had even less incremental validity when used alongside achievement tests and student high school class rank. In fact, high school rank was determined to be the best single predictor of success in college, as measured by students' cumulative grade point average. Scores for minority students (i.e., African-Americans and Hispanics) were excluded from the main analysis due to under-representation. Interestingly, however, Baron and Norman reported that when the under-represented groups were included in the analysis, SAT scores statistically significantly predicted academic performance, but when indicator variables for minority students were included in the regression model, or when these groups were analyzed separately, the regression coefficient was again not statistically significant.

On the other hand, researchers have found standardized tests to be good predictors of students' academic achievement. For example, Weller, Schnittjer, and Tuten (1992) found significantly large correlations between academic performance at the 9th- and 10th-grade levels and the scores on the Metropolitan Readiness Test (MRT) that was given to students entering the first grade. The authors concluded that the MRT had the potential to assist in predicting children's academic performance through Grade 10.

In a similar study, Qualls and Ansley (1995) investigated the relationship between students' scores on the Iowa Tests of Basic Skills (ITBS) and the Iowa Tests of Educational Development (ITED) at the elementary and middle school levels and academic success in high school. The authors indicated a significant relationship between scores on the ITBS and on the ITED at the 3rd- through 12th-grade levels and 1st-year college grade point average (GPA). Qualls and Ansley, on the other hand, concluded that the merits of using the ITBS and ITED as individual predictors of achievement at the post-secondary level appeared to be limited due to the fact that the sample did not include students from the lower end of the achievement continuum.

An earlier inquiry undertaken by Nolan and Jacobson (1972) found standardized tests administered in earlier grades to be a good predictor of students' achievement in high school. Their study investigated the relationship between course grades for ninth-grade students and the results of both the California Comprehensive Test of Basic Skills (CTBS) and the IQ scale of the California Short-Form Test of Mental Maturity (CTMM) for the same group of students at the eighth-grade level. According to Nolan and Jacobson, the CTBS was a better predictor of grades in ninth-grade English and Math courses than were scores on the scholastic aptitude or general intelligence tests.

Reading Achievement as a Predictor of Students' Competencies

Many researchers have investigated factors that influence academic achievement. For example, Weller and Weller (1997) concluded that high school students' academic abilities can be directly linked to their reading achievement. According to Espin and Deno (1993), the acquisition of reading comprehension skills was of primary importance at the secondary level because students were expected to use their reading skills to acquire content knowledge.

Generally, students with poor reading achievement have experienced difficulties in later grades. Syropoulos (1996) noted that students who were weak in basic skills at the middle school level had increased difficulties in high school. According to Ciborowski (1995), students who experienced problems reading their textbooks also encountered problems learning to read in early grades. Similarly, Simner and Barns (1991) found that students who experienced difficulty mastering the first-grade reading curriculum experienced academic problems at the secondary level.

Hanson and Farrell (1995) evaluated the long-term effects on reading outcomes among high school seniors as a function of the degree to which they learned to read in kindergarten. In this investigation, educational histories and current reading proficiencies of 3,959 high school seniors from 24 school districts in 10 states were examined to determine the effects of the Beginning Reading Program (BRP), a formal kindergarten reading program, on academic success at the high school level. Although the inquiry included Kindergarten students from all backgrounds, those from at-risk backgrounds were over-represented. Results of the study revealed significantly higher scores on all measures of reading competency for students who received reading instruction in kindergarten than for those who did not receive such instruction. Findings from the study were consistent across districts and schools, as well as across ethnic, gender, and social class groups.

Summary

Standardized tests continue to be the most popular mechanisms used to measure academic achievement. However, both the positive and negative aspects must be considered when using tests to make educational decisions regarding students. The bulk of the literature tends to indicate that standardized tests, as well as academic success in early grades are predictors of success in subsequent grade levels. In conclusion, performance on exit exams has a significant impact on high school students' ability to meet state graduation requirements. Because of the numerous factors that influence a student's ability to earn an acceptable score, it is imperative to investigate factors which predict students' performance levels on high school competency tests. Knowledge of these factors could lead to effective interventions being developed.

Methods

Participants

The convenient sample consisted of 102 students in the 1999 senior class of a high school in the state of Georgia. This sample size was selected via an *a priori* power analysis because it provided an acceptable statistical power (i.e., .79) for detecting a moderate correlation ($r = .30$) at the Bonferroni-adjusted .01 level of significance (i.e., maintaining an .05 significance level for inferences involving the 5 subtests of the GHSGT) (Erdfelder, Faul, & Buchner, 1996). The sample consisted of 78 African-American students, 22 Caucasian-American students, 1 Asian-American student, and 1 Hispanic student. A slight majority (52%) of the sample was female. Only students who took both the eighth-grade ITBS test and the GHSGT were included in the study.

Instruments

The Reading section of the ITBS (i.e., Form M, Level 14) and the Language Arts, Writing, Mathematics, Science, and Social Studies portions of the GHSGT were used to collect data for the present investigation. The Reading section contains two parts, reading comprehension and vocabulary. Students are given a total of 55 minutes to complete both sections, which are presented in a multiple-choice format. All ITBS tests are administered by professional staff members who have been oriented as to the testing procedures and to the use of the testing material. Instructions for administering the tests are included in the testing manual and must be strictly followed by the test administrator (Hoover, Hieronymus, Frisbie, & Dunbar, 1996). Brookhart (1998) reported that content-related validity on the ITBS has been provided through its evidence of alignment with texts, curriculum materials, and skills valued by school districts. Additionally, ITBS scores were reported by Brookhart as providing valid measures of basic academic skills. Brookhart further reported scores on the ITBS that yielded high reliability coefficients, as measured by KR20 (i.e., around .90).

The GHSGT was designed to measure achievement in the areas of Language Arts, Mathematics, Science, Social Studies, and Writing. All subscales, with the exception of the Writing test, are presented in a multiple-choice format. A minimum scaled score of 500 is needed to pass each subtest of the GHSGT. To receive an acceptable score on the Writing portion, students must write a persuasive essay on an assigned topic. Students are allowed 80 minutes for the Science and Social Studies tests and 60 minutes for the Language Arts and Math tests. The Social Studies portion is the longest, consisting of 90 items; Science has 80 items, Math has 70 items, and Language Arts has 60 items (Georgia Department of Education, 1991). In their 1997 report of the reliability and validity of scores generated by the GHSGT, Bunch and Klaric documented reliability estimates (i.e., KR20) for Language Arts, Mathematics, Science, and Social Studies for Spring 1994 to Winter 1996. For Language Arts these estimates ranged from .80 to .90; in Mathematics from .80 to .93; in Science from .90 to .92; and in Social Studies from .80 to .94. According to Bunch and Klaric, correlations of .45 or higher documented between course grades and scores on the English/Language Arts, Mathematics, Science, and Social Studies subscales provided evidence of construct-related validity.

Procedure

Participants were administered the ITBS during the spring of their eighth-grade year, and the GHSGT during the spring of their junior year. Students were placed into two groups based on whether they passed or failed each component of the GHSGT on the first administration. Eighth-grade ITBS Reading Normal Curve Equivalent (NCE) scores for each group were compared.

Results

Table 1 presents the intercorrelations for scores on the five subtests of the GHSGT. These correlations are presented for descriptive purposes only. That is, because the purpose of the study was to examine the relationship between the eighth-grade ITBS reading scores and the scores on the dimensions of the GHSGT, this correlation matrix was not interpreted.

Table 1
Intercorrelations for Scores on the Five Subtests of the GHSGT for the Full Sample (n = 102)[1]

Sample	Writing	Language Arts	Mathematics	Social Studies
Language Arts	.74			
Mathematics	.71	.82		
Social Studies	.72	.79	.80	
Science	.63	.78	.78	.81

[1] No levels of statistical significance are reported because these correlations did not represent the hypotheses of interest.

An examination of the skewness and kurtosis coefficients, as well as the histogram plots, indicated that the eighth-grade ITBS reading scores and GHSGT scores were approximately normal. In particular, all skewness and kurtosis coefficients fell within the normal range, justifying use of parametric statistics (Onwuegbuzie & Daniel, in press-a).

Pearson's product-moment correlation coefficient was calculated to determine the relationship between scores on the eighth-grade ITBS Reading test and the five subtests of the GHSGT (i.e., Writing, Language Arts, Math, Social Studies, and Science). Because five correlations were tested, the Bonferroni technique (Huck, 2000) was used to adjust the alpha level to .01. As reported in Table 2, all correlations were positive and statistically significant ($p < .01$).

Table 2
Correlations Between ITBS Reading Scores and Scores on Each Subtest of the GHSGT for the Full Sample (n = 102) and by Ethnicity and Gender

Sample	*n*	Writing	Language Arts	Mathematics	Social Studies	Science
Total	102	.69	.80	.80	.86	.80
Caucasian-American	22	.55	.83	.77	.88	.79
Minority	80	.67	.75	.78	.82	.76
Male	50	.53	.80	.68	.80	.78
Female	52	.77	.80	.87	.91	.82

A measure of shared variance was obtained by squaring the correlations. The strongest relationships were found between Reading and Social Studies, with 74% of the variance being shared. Moderate relationships were found between eighth-grade Reading scores and Language Arts, Math, and Science, with each pair of variables sharing 64% of the variance. The correlation between Reading and Writing represented 48% of shared variance. Using Cohen's (1988) criteria, all correlations represent large effect sizes. Thus, eighth-grade ITBS reading scores were strong predictors of student scores on the GHSGT.

Using Meng, Rosenthal, and Rubin's (1992) method for comparing correlated correlation coefficients, and applying the Bonferroni adjustment, revealed that both the correlation between Reading and Social Studies (the largest correlation) and between Reading and Language were statistically significantly higher than the correlation between Reading and Writing (the smallest correlation).

The correlations between the eighth-grade ITBS reading scores and each of the GHSGT subtests were disaggregated to determine whether there were gender differences and/or ethnic differences in the strength of these relationships (see Table 1). Specifically, a series of Fisher's *z*-tests (Onwuegbuzie & Daniel, in press-a) was conducted to compare the correlations pertaining to Caucasian-American students to those pertaining to minority students (i.e., African-American, Asian-American, and Hispanic). With respect to ethnicity, no statistically significant difference emerged between Caucasian-American and minority students for the bivariate relationships involving Writing (Fisher's $z = -0.73$), Language (Fisher's $z = 0.82$), Mathematics (Fisher's $z = -0.09$), Social Studies (Fisher's $z = 0.84$), and Science (Fisher's $z = 0.29$).

With regard to gender, although no difference was found between males and females for the correlations involving Language (Fisher's $z = 0.00$) and Science (Fisher's $z = -0.55$), statistically significant differences emerged for correlations pertaining to Writing (Fisher's $z = -2.11$), Mathematics (Fisher's $z = -2.47$), and Social Studies (Fisher's $z = -2.10$). Specifically, the relationship between eighth-grade ITBS reading scores and the GHSGT Writing scores, between eighth-grade ITBS reading scores and the GHSGT Mathematics scores, and between eighth-grade ITBS reading scores and the GHSGT Social Studies scores, were statistically significantly stronger for females than for males. The effect sizes corresponding to these statistically significant differences (i.e., differences in Fisher's *z* values) were .43, .50, and .43, respectively, suggesting moderate to large effects (Cohen, 1988).

Finally, an independent samples *t*-test was conducted for each GHSGT subtest to determine whether students who passed a subtest of the GHSGT (i.e., obtained a scaled score of 500 points or more) had attained significantly higher eighth-grade reading scores than did students who failed that particular subtest. Here, students' performance on each portion of the GHSGT (i.e., pass vs. fail) served as the independent variables, whereas eighth-grade ITBS scores were treated as the dependent variable. Although the classification of the independent and dependent variables was not consistent with the temporal sequence (i.e., the ITBS scores were obtained before the GHSGT scores), it is justified because *t*-tests, like all other members of the general linear model, represent correlational analyses, and thus do not determine the causal nature of relationships (Onwuegbuzie & Daniel, in press-b). Indeed, in discriminant analyses, the categorical dependent variable (e.g., gender) typically occurs before the predictor variables.

For all statistically significant differences, effect sizes (i.e., Cohen's *d*) were calculated by dividing the mean differences by the pooled standard deviations (Cohen, 1988). The results of these *t*-tests are presented in Table 3. It can be seen from this table that, after applying the Bonferroni adjustment: (1) students who passed the Writing subtest of the GHSGT had statistically significantly higher eighth-grade reading scores than did those who failed the Writing subtest; (2) students who passed the Language subtest of the GHSGT had statistically significantly higher eighth-grade reading scores than did those who failed the Language subtest; (3) students who passed the Mathematics subtest of the GHSGT had statistically significantly higher eighth-grade reading scores than did those who failed the Mathematics subtest; (4) students who passed the Social Studies subtest of the GHSGT had statistically significantly higher eighth-grade reading scores than did those who failed the Social Studies subtest; and (5) students who passed the Science subtest of the GHSGT had statistically significantly higher eighth-grade reading scores than did those who failed the Science subtest. Using Cohen's (1988) criteria, the effect sizes corresponding to these five differences were extremely large.

Discussion

This study investigated the relationship between reading scores on the ITBS and each of the subscales of the GHSGT (i.e., Writing, Language Arts, Math, Social Studies and Science) for eighth-grade students. The correlational analyses indicated statistically significant relationships between eighth-grade ITBS Reading scores and each subscale of the GHSGT. The amount of variance shared between each pair of variables was particularly high, suggesting that ITBS reading scores at the eighth grade are significant predictors of performance at the high school level, as measured by all five subtests of the GHSGT. Thus, these findings provide evidence to support the validity of using ITBS Reading scores as a predictor of student scores on the GHSGT.

A comparison of the five bivariate relationships, using Meng et al.'s (1992) procedure, indicated that although ITBS reading scores were reliable predictors of all GHSGT subtests, the former was an even better predictor of the Social Studies and Language subtests than of the Writing subtest. Future research should investigate why reading performance at the eighth-grade level had greater predictive power for social studies and language than for writing.

Interestingly, no significant ethnic differences were found with respect to the bivariate relationships. That is, eighth-grade ITBS reading scores had similar predictive power across ethnic lines. However, these findings should be interpreted with caution because of the relatively small proportion of Caucasian-American students in the sample.

On the other hand, gender differences emerged with regard to some of the bivariate relationships. Specifically, ITBS reading scores were a better predictor of the Writing, Mathematics, and Social Studies subtests of the GHSGT for females than for males. These results suggest that reading achievement at the eighth-grade level may be even more important for girls than for boys. Unfortunately, it is beyond the scope of the present investigation to determine why this might be the case. Thus, this should be a focus of future inquiries.

A series of independent *t* tests revealed that students who passed a GHSGT subtest had obtained statistically significantly higher reading performance levels at the eighth grade than did their failing counterparts. Each mean difference represented a large effect size, thereby indicating that reading achievement has a substantial impact on the GHSGT pass rate. Moreover, students who failed a portion of the GHSGT typically scored in the lowest quartile of the distribution of eighth-grade ITBS reading scores. This finding suggests that eighth-grade students at the bottom quartile in reading achievement could be considered at-risk for future academic failure, and perhaps, be the target of academic-related interventions.

Table 3
Performance Differences on the Eighth-Grade ITBS Reading Test Between Students Who Passed and Failed Each Subtest of the GHSGT

	Passing Students			Failing Students				
Variable	M	SD	N	M	SD	N	t-Value	Cohen's *d* Effect Size
Writing	43.80	21.44	88	15.31	21.44	14	5.74*	1.37
Language Arts	44.35	20.94	88	10.79	12.26	14	8.47*	1.68
Mathematics	49.44	19.71	68	20.35	16.06	34	7.98*	1.57
Social Studies	54.01	19.42	55	23.04	14.02	47	9.32*	1.81
Science	53.50	20.59	54	24.27	14.24	48	8.41*	1.63

* $p < .001$

In any case, findings from the present study support the hypotheses that a relationship exists between low reading scores on the ITBS and failing scores for each sub-scale of the GHSGT. Indeed, all five hypotheses were supported. The current results are consistent with Weller and Weller's (1997) contention that strength in reading comprehension is a predictor of academic failure in high school. The fact that students who failed the subscales of the GHSGT had significantly lower mean reading scores than did students who passed each test provides evidence to support conclusions from Weller et al. (1992) and Qualls and Ansley (1995) that standardized test results in early grades can be beneficial in predicting academic performance in subsequent grades. Results from this investigation also are consistent with the findings of Cibrowski (1995), Simner and Barns (1991), and Syropoulos (1996), who traced the origin of low academic performance of high school students to insufficient reading skills in early grades.

Implications of Findings

Although all results from the study were found to be statistically significant, the reader must be reminded of certain limitations which may threaten the internal and external validity of the findings. First, because more than a two-year period existed between the administration of the ITBS Reading Test and the GHSGT, the higher mean scores among the students who passed the test could be attributed to the passage of time. That is, maturation was a possible threat to internal validity (Gay & Airasian, 2000).

Additionally, generalizations to other populations are limited by the fact that the sample included a relatively small number of 12th-grade students from a rural community in south Georgia. Thus, replications of this study are needed to determine the reliability of the present findings. Moreover, because the use of the GHSGT as a measure of high school students' competencies is recently new, future research is needed to determine if similar results are present in high schools throughout Georgia.

Recommendations for Future Practice

Although eighth-grade reading scores were a significant predictor of performance on the GHSGT, it is possible that reading achievement at earlier grades is an even better predictor of high school performance. Thus, future research should examine trends in reading performance over time, and how these trends relate to the GHSGT. Such investigations could help to determine the grade at which interventions are most likely to be effective in improving high school achievement.

In any case, the present findings indicate that the ITBS may be an appropriate screening instrument to identify students who are at-risk of failure before they enter high school. Moreover, the current results suggest that student performance on the GHSGT might be increased by providing at-risk students with appropriate reading interventions at or before the eighth grade. School systems also should explore the possibility of implementing a remedial reading program at the ninth-grade level. Evidence presented in this study suggests that such programs might be effective in improving student performance on the GHSGT and, hence, increasing the graduation rates among high school seniors.

References

Baron, J., & Norman, M. F. (1992). SATs, achievement tests, and high-school class rank as predictors of college performance. *Educational and Psycho-logical Measurement, 52*, 1047-1054.

Brent, G., & DiObilda, N. (1993). Effects of curriculum alignment versus direct instruction on urban children. *Journal of Educational Research, 86*, 333-338.

Brookhart, S. (1998). Iowa Tests of Basic Skills, Forms K, L, and M. In Impara, J.C., Murphy, L. L., & Plake, B. S. (Eds.*), 13th Mental Measurements Yearbook* (pp. 540-542). Lincoln, NE: University of Nebraska.

Bunch, M. B., & Klaric, J. S. (1997). *Georgia High School Graduations Tests: Reliability and Validity* (Measurement Incorporated).

Charlesworth, R., Fleege, P. O., & Weitman, C. J. (1994). Research on the effects of group standardized testing on instruction, pupils, and teachers: New directions for policy. *Early Education and Development, 5*, 195-209.

Ciborowski, J. (1995). Using textbooks with students who cannot read them. *Remedial and Special Education, 16*, 90-101.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: John Wiley.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*, 1-11.

Espin, C., & Deno, S. L. (1993). Performance in reading from content area text as indicators of achievement. *Remedial and Special Education, 14*(6), 47-59.

Gandal, M. (1995). *Making standards matter: A fifty-state report on efforts to raise academic standards.* Washington, DC: American Federation of Teachers.

Gay, L.R., & Airasian, P.W. (2000). *Educational research: Competencies for analysis and application* (6th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Georgia Department of Education. (1991). *Research, Evaluation & testing Student Learning & Assessment* (Rule Number 160-3-1-07). Atlanta, GA.

Georgia Department of Education. (2000). *Georgia High School Graduation Tests (GHSGT)*. Retrieved March 29, 2000 from the World Wide Web: http://www.k12.ga.us/sla/ret/ghsgtabout.html

Griffin, B.W., & Heidorn, M.H. (1996). An examination of the relationship between minimum competency test performance and dropping out of high school. *Educational Evaluation and Policy Analysis, 18*, 243-252.

Hanson, R. A., & Farrell, D. (1995). The long-term effects on high school seniors of learning to read in kindergarten. *Reading Research Quarterly, 30*, 908-933.

Herman, J. L., Abedi, J., & Golan, S. (1994). Assessing the effects of standardized testing on schools. *Educational and Psychological Measurement, 54*, 471-482.

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1996). *ITBS directions*. Itasca, IL: Riverside Publishing Company.

Huck, S.W. (2000). *Reading statistics and research* (3rd ed.). New York: Addison Wesley Longman.

Jacobson, L. (1998, June 24). State graduation tests raise questions, stakes. *Education Week, 17*(4), pp. 22, 26.

Lessitz, R. W. (1997). Statewide performance assessment: Continuity, context, and concerns. *Contemporary Education 69*(1), 15-19.

Levande, D. (1993). Standardized reading tests: Concerns, limitations, and alternatives. *Reading Improvement, 30*(2), 125-127.

Lomax, R. G., West, M. M., Harmon, M. C. Viator, K. A., & Maddaus, G. F. (1995). The impact of mandated standardized testing on minority students. *Journal of Negro Education, 64*, 171-185.

Meng, X-L., Rosenthal, R., & Rubin, D.B. (1992). Comparing correlation coefficients. *Psychological Bulletin, 111*, 172-175.

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Transforming testing in America*. Chestnut Hill, MA: Author.

Nolan, J. S., & Jacobson, J. (1972). The California Comprehensive Test of Basic Skills: A predictor of success. *Educational and Psychological Measurement, 32*, 451-452.

Norton, S.M., & Park, H. (1996, November). *Relationships between test preparation and academic performance on a statewide high school exit examination*. Paper presented at the Annual Meeting of the Mid South Educational Research Association, Tuscaloosa, AL. (ERIC Document Reproduction Services No.: ED 406 426)

Onwuegbuzie, A.J., & Daniel, L.G. (in press-a). Uses and misuses of the correlation coefficient. *RESEARCH IN THE SCHOOLS, 9*(1).

Onwuegbuzie, A.J., & Daniel, L.G. (in press–b). Typology of analytical and interpretational errors in quantitative and qualitative educational research. *Current Issues in Education*.

Qualls, A. L., & Ansley, T.N. (1995). The predictive relationship of ITBS and ITED to measures of academic success. *Educational and Psychological Measurement, 55*, 485-498.

Sacks, P. (1997). Standardized testing. *Change, 29*(2), 24-31.

Schmeiser, C.B., & Ferguson, R.L. (1979). Effects of culture-based test materials on performance of black and white examinees. *Educational Horizons, 58*(1), 53-58.

Simner, M. L., & Barns, M.J. (1991). Relationship between first–grade marks and the high school dropout problem. *Journal of School Psychology, 29*, 331-335.

Strong, S., & Sexton, L.C. (1996). Kentucky performance assessment of reading: Valid. *Contemporary Education, 67*(2), 102-106.

Syropoulos, M. (1996). A support program for at-risk ninth graders: Evaluation of the ninth grade project at Denby High School. *ERS Spectrum, 14*(3), 17-26.

Waltman, K.K. (1997). Using performance standards to link statewide achievement results to NAEP. *Journal of Educational Measurement, 34*, 101-121.

Weller, L. D. & Weller, S. J. (1997). Using Deming's Continuous Improvement Model to improve reading. *NASSP Bulletin, 81*, 78-85.

Weller, L. D., Schnittjer, C. J., & Tuten, B. A. (1992). Predicting achievement in grades three through ten using the Metropolitan Readiness Test. *Journal of Research in Childhood Education, 6*, 121-130.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 2, 11-19

From Texas to Florida: Email Peer Coaching Dialogues Among Preservice Teachers

Susanne I. Lapp
Florida Atlantic University

Educators in Florida and Texas have made great strides to create successful learning opportunities for elementary students. To assist in these efforts, teacher educators have carefully scrutinized the educational experiences of preservice teachers, looking specifically at preservice teachers' application of pedagogical knowledge to the learning environment. Teacher educators have created more meaningful and supportive experiences for preservice teachers prior to and during student teaching. The following paper describes the efforts made to electronically link preservice students from educational institutions in Texas and Florida for the purpose of sharing and exchanging new, dynamic ideas and activities in the classroom environment. Results suggest that email peer coaching dialogues made preservice teachers more aware of the challenges faced by educational professionals as they attempt to improve the learning opportunities of elementary students. Peer coaching dialogues provide preservice teachers with an opportunity to share new and innovative ideas to make the learning environment more stimulating and exciting for elementary students. Finally, through email dialoging, participants are able to establish vital communicative links with other preservice teachers, thus easing their transition from university student to professional educator.

One of the most challenging issues confronting preservice teachers upon entering the teaching profession is to combine their knowledge of content and methodology with classroom practice. Once preservice teachers graduate from teacher preparation programs and transition into classrooms, they are expected to modify and re-evaluate content and instructional style in an effort to create a positive learning environment for their students. As new teachers integrate content and instructional modifications, they typically develop an understanding of what works best in the classroom and begin to identify themselves as professional educators (Borich, 1995).

Hall (1968) and Kerr, VonGlinow, and Schriesheim (1977) researched professional identity development and have suggested that it occurs in two stages: structural and attitudinal. The structural stage occurs externally as the new teacher acquires the requisite skills, certifications and degrees necessary for entry into the teaching profession. Once the individual successfully completes the structural stage, it is assumed that he/she is prepared for the second stage in his/her professional identity development, the attitudinal stage. In the attitudinal stage, the teacher has internalized the skills and pedagogical knowledge learned as a student and is able to apply this knowledge in the classroom. Successfully passing through both the structural and attitudinal stages helps the individual develop a professional identity as an educator (Brott & Kajs, 2000).

Unfortunately, the transition from the structural to the attitudinal stage is difficult for many preservice teachers and has been the focus of research literature in recent years (Kleinsasser, 1988; Richards, 1992; Riley, 1998; Snow, Burns & Griffin, 1998; Spatig, 1994) with most studies addressing preservice teachers' lack of preparation as they transition from the role of student to teacher. The transitional phase is often characterized by ambivalent feelings, insecurity and contradictions. The preservice teacher is neither a full-fledged student, nor teacher, and as such does not possess a professional identity as an educator (Cordeiro & Smith-Sloan, 1995; Furlong, 1997) .

Peer dialogues

To strengthen new teachers' professional identity development, new and innovative techniques including peer coaching and peer dialogues, have been developed (Cole, 1995; Leggett & Hoyle, 1987; McAllister & Neubert, 1995; Richards, 1992). Peer coaching encourages professional dialogue among preservice teachers. During peer coaching, participants share stories and experiences with other preservice teachers in an environment that is supportive and nurturing. Participants discuss innovative educational ideas and strategies and provide suggestions for overcoming obstacles to effective

---

Susanne I. Lapp is an Assistant Professor of Reading and Language at Florida Atlantic University. Please address correspondence regarding this article to Susanne I. Lapp, Reading and Language Arts, Florida Atlantic University, Davie, FL 33314 or by e-mail to slapp@fau.edu.

instruction (Bercik, 1992). Through discussion with peers, preservice teachers receive valuable feedback which helps to strengthen their knowledge in the classroom and their professional identity as educators.

Reflective Strategies in Peer Coaching

Brott and Kajs (2000) suggest that a critical component for successful peer coaching involves effective communication and dialogue. Peers must be flexible and collaborative as they share educational experiences and provide each other with beneficial feedback. Brott and Kajs further divide peer coaching into three distinct categories:

1. attending and listening
2. reflecting and clarifying
3. challenging and confronting

Attending and Listening: In attending and listening, the participants must first become acquainted with each other before they can discuss educational issues. At this stage, participants become aware of and appreciate the unique nature of their peers' individual behavior. To effectively gauge their peers' behavior, participants must begin communication by learning as much as they can about their peers likes and dislikes, their cultural backgrounds, and any specific geographical influences which may influence their attitudes and beliefs. In addition, peer coaching participants must be able to cultivate deeper levels of communication through active listening. Active listening encourages participants to freely share their thoughts, feelings and actions without fear of judgment or negative retribution.

Reflecting and Clarifying: Peer coaching participants find that as they actively listen to each other, they begin to paraphrase and mirror their peers' experiences. Peers reflect on similar experiences and provide insight on how they may have dealt with similar situations. This spirit of empathy and understanding among peers provides a foundation for the second level of peer coaching: reflecting and clarifying. At this stage, participants are already familiar with their partners' backgrounds and beliefs, and this information serves to guide more in-depth discussion on educational issues. Participants are asked to reflect or clarify their actions or reactions to classroom situations, based on their personal beliefs about education and educational theory. By encouraging their partners to reassess classroom events, peers are able to stimulate discussion and encourage deeper exploration of their evolving professional identities.

Challenging and Confronting: The final category in peer coaching focuses on challenging and confronting. The purpose of this stage is to encourage participants to become actively involved in specific educational issues. Peers are encouraged to face classroom challenges and confrontations instead of trying to avoid them. Bercik (1992) found that many preservice teachers were concerned with specific educational issues and societal expectations placed on them as future teachers. Specifically, preservice teachers were concerned whether they were sufficiently prepared to deal with a number of classroom challenges including: classroom management and discipline, understanding the needs of diverse students, and effectively blending theory and practice.

Additional Challenges

In states like Florida and Texas, students must deal with similar concerns, including minority and English as a second language (ESL) student education. Florida and Texas have some of the highest percentages of minority student enrollment in the United States (Aleman, 1993; Sietsma & Bose, 1995; Trueba, 1998) most of whom are Hispanic and in the Kindergarten through sixth grades (Denton, 1993; Ngyuen, 1987). To help these children learn to read, write, and speak English, legislative decisions in both states have mandated English language training and bilingual education. The META Consent Decree enacted by the state of Florida in 1990 requires that all elementary public school educators, who are the main providers of English instruction and have ESL students in their classrooms, must obtain ESOL (English for Speakers of Other Languages) endorsement for their teaching certification. Vigorous debate over ESOL legislation continues in Florida and Texas.

In summary, encouraging preservice teachers to take a proactive stance when dealing with educational issues and societal expectations prepares them for the realities of the classroom and encourages them to become more responsible for selecting the most appropriate method of instruction. By engaging in peer coaching activities, preservice teachers have the opportunity for reflective thinking and professional growth as they begin to apply their developing pedagogical knowledge and new teaching techniques while in the classroom (Snow et al., 1998).

Purpose of the Study

As a teacher educator in Florida, I was curious to see whether educational issues or societal expectations impacted preservice teachers' developing professional identity in the field of education. As a result, I created a project which addressed two focal questions:

1. What educational issues receive the greatest interest among preservice teachers as they engage in peer dialogue discussions?
2. What solutions are generated by preservice teachers when confronted with various educational issues?

To answer these questions, preservice teachers from one teacher preparation program in Florida were electronically linked via email to another teacher preparation program in Texas. The email messages of the participants were collected and analyzed. The preservice teachers' responses are discussed below. The report begins with a brief discussion of the background of the project and a description of the data collection and data analysis procedures.

Methodology

Background of the project

The study took place in two separate university sites in the southern region of the United States. The first university site was located in a large, metropolitan area along the southeastern coast of Florida. The second university site was located in a small, rural town along the Texas-Mexico border.

Participants

To solicit participation in the study, preservice elementary education majors from two separate universities in Florida and Texas, who were taking a Reading Methods course, were informed about the project. The preservice teachers represented a nontraditional population. The average age of preservice teachers was 30. Fifty percent of these preservice teachers were married with children, and the remainder were single or divorced. Seventy-seven percent of the preservice teachers worked full time while attending classes (Campus Update, 1999). Preservice teachers from Texas were first generation Mexican-American, while Florida preservice teachers were African American and second and third generation Caribbean and European Americans.

All of the preservice teachers were in their last semester before student teaching and were taking their final sequence of upper division education courses at their respective universities. Both educational programs in Texas and Florida were experimental in design and provided preservice teachers with early exposure to elementary classrooms. Schools in Florida and Texas attempted to reform their curriculum and education (Archbald, 1994) with a concentration on improved academics, teacher professionalism, enhanced financial support and educational technology (Firestone et al., 1990). Florida preservice teachers participated in classroom related instructional activities beginning in their junior year of high school. Their presence in the elementary classroom increased throughout their college years until they assumed complete control over the classroom during their traditional semester of student teaching at the end of their senior year in college. Texas preservice teachers also spent a considerable amount of time in elementary classrooms, where they participated in field-based teaching. All of their education methods courses were taught on elementary campuses throughout the south Texas region and provided preservice teachers with easy access to elementary age students.

In addition to early classroom experience, both teacher preparation programs infused their courses with educational technology. Preservice teachers had easy access to computer centers which were conveniently located at their universities and on several elementary campuses. Preservice teachers were required to demonstrate their skills in educational technology by incorporating multimedia authoring packages such as Hyperstudio (Saulpaul & Badiner, 1997) and web-based activities including the internet and email, all of which they used during their fieldwork experiences in the schools. During the first week of the semester, preservice teachers were informed about the email peer coaching project.

Data Collection

Preservice teachers were told that participation in the email peer coaching dialogue project would provide them with an opportunity to share ideas and interact with preservice teachers from another university through email exchanges. They were also aware that their instructors would collect data on their email exchanges. The instructors informed the preservice teachers that the email peer dialogue project was not a course requirement and would be undertaken on a voluntary basis. Preservice teachers were reassured that they would not be penalized for refusing to participate in the program and were free to discontinue participation at any time during the project. They were informed that the project would last for 14 weeks and that they would be encouraged to communicate with their partners frequently throughout the semester. Preservice teachers who indicated an interest were asked to sign consent forms. As a result, nine preservice teachers from Texas and nine preservice teachers from Florida volunteered to participate in the project. Each preservice teacher from Florida was paired with a preservice partner from Texas.

The data sources used in this study consisted of over 250 email messages. These email messages were collected throughout the duration of the 14 week study. Each of the preservice teachers' email accounts were configured so that every email entry sent between partners was also copied and forwarded to the instructors' accounts where an archive file was created so that email data could be stored for future analysis.

Data Analysis

The goal of the study was to determine which educational issues sparked the greatest interest among preservice teachers as they engaged in on-line, peer coaching dialogue discussions. During the data analysis stage of the project, the Reading Methods instructors at the Texas and Florida sites read and reread the entire set of email data searching for (a) answers to the focal research questions and (b) patterns of email exchanges which emerged.

In order to provide consistency in counting the number of email topics elicited by the participants in the study and provide a method of coding the email messages, T-unit analysis was employed (Hunt, 1965). Hunt defined T-units as whole pieces of writing which were separated into units of information. Each unit consisted of one main clause with any additional subordinate clauses attached to it. The T-unit was the shortest grammatically allowable sentence into which a theme could be segmented without creating a sentence fragment. The following passage provides an example of how a sample passage from one of the email messages was divided into T-units. A slanted line in the example identifies each T-unit.

> Well, I think we know a great deal about each other./ I've learned a lot about the South Texas 'Valley'/ and we have had a chance to share some really neat stories./ Now lets get down and talk about some of these school issues/ and what we will face when we get out there./

After each email exchange was divided into T-units, they were coded and sorted into topics to provide a level of categorization. The instructors randomly selected eight sections of the coded email data sources in order to check the reliability of the analysis. The results indicated a strong agreement (88%) between the two raters, indicating an acceptably high level of reliability (Miles & Huberman, 1994). Data analysis revealed several strong patterns or stages among the email messages. The topics within these stages closely resembled the three categories: attending and listening, reflecting and clarifying, and confronting and challenging, which were used in the Brott and Kajs (2000) study on peer coaching. The instructors decided to use these established categories as a useful means of categorizing the email exchanges for the present study. Thus, the email messages were categorized as attending and listening, reflecting and clarifying, and confronting and challenging and are discussed below.

Attending and Listening. All nine email partnerships initiated their email peer coaching exchanges with an initial warm-up stage which was referred to as the attending and listening stage. Participants used this stage as an opportunity to get to know each other. Peers shared important personal aspects of their lives including information about themselves, their hometowns, educational programs, and peer coaching expectations. These messages helped to set the tone for future dialogues on professional identity development.

Diane from Florida (FL) described life in South Florida to her partner Norma from Texas (TX), " Life can be really crazy here. There are so many people moving into Florida everyday and that's why we have school overcrowding. Some come for the weather and all the Florida excitement and others come for the jobs." Norma quickly responded to Diane's message, " I know about 'the Florida life' from TV, but my hometown is a bit more laid back. Our nearest big city is San Antonio, but that town is also considered to be in South Texas."

Students also used the attending and listening stage to exchange stories about cultural traditions. Norma (TX) described a typical Sunday dinner with her family:

> I thought you might be interested in one of our typical Mexican traditions. It may sound a little different from what you might be used to. On Sunday's my family gets together for dinner which is prepared in traditional Mexican style. The men are served first, then the children and finally the women.

Several days later, Diane (FL) responded with interest and appreciation, "Your story is really fascinating, but don't you get hungry having to serve everyone first? Do other families celebrate the same way?" Email exchanged during the attending and listening stage of communication was personal and provided participants with a sense of familiarity and comfort. Sharing life experiences created an element of trust and emotional support among the peer coaching participants.

Other participants, during the attending and listening stage were also interested in sharing their experiences while at the university. Shannon (FL) addressed some of her struggles while attending the university, "It's hard for me to go to school and work. We all have alot [*sic*] of obligations in the program. Studying, attending classes, observations (in the school). Then I have to go home and take care of the kids." Shannon's experiences were similar to many of the preservice teachers in both teacher preparation programs. Preservice teachers were expected to balance the professional demands of becoming an educator with the personal demands of raising a family. Shannon's partner, Tammy (TX) agreed with her comments, but added another element to the discussion, "Yes, we do have to make many sacrifices. I would also like to have fun with my friends outside of school, but we have so many tasks and pressures and we must prepare ourselves to pass all the teacher exams." Several email exchanges reflected preservice teachers' fear of exit proficiency exams. As part of Florida and Texas' educational reform movement, preservice teachers were required to pass a variety of performance measures including successfully taking and passing exit proficiency exams prior to entering student teaching (Laitsch, 1998).

Since the peer mentoring project was conducted online and all communication occurred via email, participants felt that it was critical to establish rules for email communication. Most students began their communication efforts very enthusiastically. In an email messages to Maria (FL), Elsa (TX) wrote:

> This is going to be an exciting experience to talk with you through email and we must work to constantly send messages back and forth. I look at this experience as being the first of many 'tech' connections throughout the country with colleagues who share the same interest in education and desire to help children become all they can be.

Occasionally, messages addressed the lack of active communication among several partners. Andrea (FL) responded:

> I am really enthusiastic about this communication with other students in Texas; however, I am becoming anxious due to the lengthy wait for a response. It is important for us to begin emailing as soon as possible. I have alot [sic] of things I need to ask my partner and I hope my anticipation of new friendships and learning will not become a disappointment.

The importance which peer coaching partners placed on their email dialogues are similar to the findings by Brott and Kajs (2000), who stressed the importance of effective dialoguing skills among peer coaching participants. As peer relationships were being built and experiences shared, participants needed to be reassured that their email partners could be trusted with this personal information.

Some partners used their educational demands and pressures from home as excuses not to send messages. Adam (FL) apologized to Thomas (TX) for his delayed response: "By the time I get home from school, I have to finish studying and then eat dinner. I get the kids ready for bed and then I forgot to send you an email - sorry!" Two of the nine partnerships remained at the attending and listening stage of communication throughout the duration of the project. These participants continued to communicate with the same frequency like the remaining seven participants, one set of exchanges per week; however, they used the email dialogues as an opportunity to revisit the same topics dealing with academic and family pressures. Although these partners did make an effort to communicate, the majority of email exchanges were not sufficient to sustain deeper conversation on issues related to education.

Clarifying and Reflecting. Seven of the partnerships managed to advance to more specific levels of peer coaching dialogues. In a message to Norma (TX), Diane (FL) appeared to set topical parameters for future email dialogues:

> Well, I think we know a great deal about each other. I've learned a lot about the South Texas 'Valley' and we have had a chance to share some really neat stories. Now lets get down and talk about some of these school issues and what we will face when we get out there.

This exchange indicates more mature levels of professional identity development. Having established a foundation of trust and commitment to continue on with the exchanges, both partners indicated a desire to concentrate on specific knowledge and skills which are required of future professional educators. Other peer coaching groups followed similar patterns. As a result, several topics emerged from the email data which highlighted preservice teachers' interest in using the dialogues to explore and gain better understanding of various educational issues such as incorporating multiculturalism and effective first and second language learning strategies in the classroom.

Diane and Norma discussed language learning strategies based on their experiences in the schools. Norma claimed:

> According to what we read in class, it is very important for students to use their first language (L1) in order to acquire second language (L2) skills. However, what I see in the schools is Spanish being used exclusively in several classrooms where the population is 100% Mexican-American, instead of trying to make the classrooms more Spanish / English bilingual.

In this example Norma has identified a problem between theory and practice. She is well aware of the benefits of bilingual education, yet she has identified a potential problem in several of the classrooms where she was working. Diane agreed with Norma and cited similar examples in Florida:

> Yes, I see the same thing happening here in Florida. I've seen several teachers give special attention to Spanish kids who have limited English communication skills but, there are so many languages represented here in Florida that many of these other kids are dropped into ESL classes where they remain for a long period of time. Teachers need to have a better solution to this problem and try and help these kids.

Diane responded to Norma's query through mirroring and clarifying. Diane supported Norma's comments by mirroring similar events which occurred in Florida. She added to the discussion by introducing problems associated with ESL programs. Diane's suggestion that teachers need assistance to improve students' learning indicates that she is trying to clarify the problem and explore potential solutions to help limited English proficient children.

Other participants were also motivated to discuss language issues and their impact on teacher effectiveness. Nora (TX) commented to her partner Beth (FL):

> I saw a teacher using Spanish to communicate and instruct students. After a while the teacher discovered that a child was not participating in the discussion. She talked with the student and learned that the child was Filipino. She reluctantly had to change the whole instructional format to accommodate the non-Spanish speaking student with English instruction.

Beth responded, "That makes me really angry. Imagine being that child. I would have been so upset! It took too long to realize the child didn't speak Spanish. It's the teacher's responsibility to introduce English into the discussion." In this example, Nora disclosed a language issue which occurred in the classroom she was observing. Beth's strong reaction to the situation suggests that she is using her personal reaction to the situation to make sense of the teachers' professional responsibilities.

Other participants struggled with similar reflection and clarification issues in their email dialogues. In an exchange between Lisa (FL) and Francis (TX), Lisa explained some difficulties she was experiencing with classroom management. Several of the students in her classroom, including two young boys from the Dominican Republic had been disruptive during a lesson. After the lesson, Lisa called the students to her desk to discuss their behavior. According to Lisa, "Every time I tried to make eye contact with them, they seemed to look at the floor. I don't think they were listening to me-how disrespectful!" Francis immediately replied to Lisa's email by offering to clarify the students' behavior. Using information from Nieto's (1996) research on multiculturalism in American schools, Francis explained that the two students may have responded to Lisa through their own cultural lenses. In Latin American countries, students will not make eye contact with a teacher as a show of respect, not disobedience. Francis gently reminded Lisa that American teachers tend to believe that a student should, "look me in the eye when I am talking to you." Lisa was intrigued by her students' reactions and how they clearly reflected Nieto's research. Lisa sent a message to Francis thanking her for clarifying the situation and making her aware of the importance of going beyond one's own cultural expectations in order to understand how students may interpret those expectations.

Confronting and Challenging. As the project progressed, the level of openness and honesty among peers intensified. At the confronting and challenging stage, email messages became more focused as preservice teachers began to assume more responsibility in the classroom. In the email messages, peers challenged and encouraged each other to confront obstacles in their pursuit of appropriate educational solutions. Many preservice students frequently incorporated their email partners' ideas and feedback while teaching reading in their field placements. Email messages increasingly suggested that peers were drawing on educational research and theory to make their lessons more effective. After observing how many elementary students had minimal English language input during classroom instruction, Hilda (TX) decided to use a Read Aloud (Routman, 1991) activity with some of the elementary students in her classroom. Afterwards, she detailed her success to her email partner Andrea (FL):

> I had the chance to try a 'read aloud' in the classroom. The kids seemed to like this approach to reading. I could tell they were working on their listening skills because a few of them actually paid more attention to me! I told my mentor teacher that our class text said that read alouds help students develop a sense of story schema and they are able to work on their prediction and response skills. I also think it's a natural way for the kids to work on their English language acquisition.

After learning of Hilda's success using the 'read aloud' strategy, Andrea was challenged to replicate Hilda's success in her own classroom:

> I took your advice about read alouds and it was great. Yesterday, I had the chance to work with a reading group and I used the multicultural text, *Friday is Papa Night* by Ruth Sonneborn. The story is about a Hispanic family who celebrate the arrival of their father who returns home after working at two jobs. In the story they have a real big party. All of a sudden, a bunch of the kids who were listening to me read, started to talk about parties and celebrations in their homes. They were so excited and wanted me to read the story over again! Read alouds really seem to motivate students to read!

In another example, Nora, Francis and Elsa from Texas shared news about gang related activity in the vicinity of the school. In her email messages, Elsa confided her fear about teaching in a dangerous environment. Elsa wrote to her partner Maria, "How can kids learn in an area filled with gang violence. Playgrounds littered with smashed bottles, and gang-related graffiti spray painted on the walls and trees." Maria listened to Elsa's cry for help, but challenged Elsa and her classmates to confront the issue and find a reasonable solution to the this problem. Maria suggested that Elsa and her classmates create a lesson plan on a social injustice issue. Maria suggested that they focus the lesson on the defined needs of the elementary students by highlighting the presence of gangs in the community. Maria and two other preservice teachers from Florida assisted their Texas peers by researching gang violence issues. They found information from the Internet, magazine articles, news clippings and personal stories related to the negative impact of gangs. They quickly emailed their Texas peers with the information and included suggestions for creating appropriate and interesting lessons on gang violence.

Elsa gratefully acknowledged the efforts made by the preservice teachers in Florida. Upon completion of the social injustice activity Elsa emailed her partner, Maria:

> Francis, Nora and I gave the kids an assignment to draw a picture of their community. (Thanks Maria for that idea!) Like we had anticipated, several students drew gang symbols with characters giving gang-related hand signals. In response to their drawings, we created a teaching/learning situation and provided these elementary students with an opportunity to see some alternatives to gang violence. We made a K-W-L Chart (Ogle, 1986) like the one we had used in our language arts class at school. Here's how it looked.
>
> Gang violence in South Texas.
> K: What they *know* about gang violence.
> W: What they *want* to learn about avoiding gang violence.
> L: What they have *learned* about gang violence.
>
> I really think the students had a chance to critically look at gang violence and the impact it makes on all our lives.

Although all email exchanges provided valuable information, those exchanged during the confronting and challenging stage provided compelling evidence that preservice teachers showed signs of developing a professional identity as educators. They became more highly skilled at directing their learning, providing effective feedback, and applying theoretical knowledge to the classroom. Email dialogues encouraged preservice teachers to become more critical consumers of educational research as their exchanges began to focus on sharing ideas and strategies to resolve student learning problems and conflicts.

Discussion and Implications

One of the major themes that emerged from the study was the level of commitment maintained by the peer coaching email participants. From the outset of the project, participants made their email communication expectations clear. As the findings from Brott and Kajs (2000) suggest, successful peer dialoging is determined by participants' effective use of communication. By advancing through the peer dialoging stages of attending and listening, clarifying and reflecting, and confronting and challenging, they were able to manifest a commitment to their partners and respect for the teaching profession. Although the email partners were stressed by pressures from home and school, the majority of participants were committed to the email project and to each other.

Analysis of the emails indicated that preservice teachers also showed a high level of commitment to the profession. The email peer coaching dialogues provided preservice teachers with an opportunity to work through their personal reactions to classroom events and societal influences. The participants were consciously aware of their future responsibilities as professional educators and showed no signs of ambivalence or insecurity as was suggested in the research by Kleinsasser (1988), Richards (1992) and Snow, Burns and Griffin (1998). Most of the preservice teachers in this project exchanged stories of success and failure as they began to apply educational research in the classroom setting. As peers moved through the stages of peer coaching, they began to share information on effective discipline strategies, desired behavioral outcomes and plausible alternatives to violence. These activities prepared them to deal with the challenges faced by classroom teachers.

Preservice teachers also found it important to voice their opinions on controversial language learning issues and became increasingly aware of the educational roadblocks that many of their future students encountered as they attempted to learn their second language, English. Instead of simply complaining about what they witnessed in the schools, preservice teachers rallied behind each other and developed innovative strategies and solutions to deal with the identified problems. In an effort to find activities which worked best in the classroom (Borich, 1995), preservice teachers shared successful literacy activities which motivated children to read and communicate in English. The email participants refused to permit instructional obstacles defeat their desire to become members of the teaching profession and believed that they could make a positive impact in the lives of children.

Another finding from the project was the positive impact of technology on the partners' developing professional identity. Communicating via email provided students with the opportunity to establish relationships with other students across the country who were willing and eager to dialogue on specific educational issues. Students had the opportunity to share stories and experiences, receive feedback on lessons, and exchange ideas on English language learning and multicultural awareness issues. Clearly, teachers who have the opportunity to use educational technology for a functional purpose will be more informed, active users of the technology in their future classrooms.

In summary, the goal of linking preservice teachers from Texas and Florida in order to strengthen and refine their professional identity as educators was achieved through this email peer coaching dialogue project. The project provided students with a real context for extending and redefining their ideas and beliefs about teaching. By communicating with peers, many of the preservice teachers moved beyond the cross-cultural boundaries, which they had previously constructed, and became aware of other cultural traditions and behaviors. Preservice teachers discovered that their cultural backgrounds might be different, but their hopes, aspirations and challenges as future teachers were similar. Through peer discussion groups, preservice teachers were able to critically evaluate their attitudes towards teaching and the teaching strategies advocated in the research literature. They had a better understanding and appreciation for the field of literacy, educational theory and computer technology. Preservice teachers gained a clearer impression of the challenges which awaited them as they entered the teaching profession and an understanding of commitment and dedication necessary to become effective teachers.

Results from this project suggest that we, as teacher educators, must respond to the needs of preservice teachers by creating opportunities for them to interact, share and evaluate ideas either through on-line or face to face discussions, so that they are able to manifest the high level of commitment and dedication necessary to become effective educators. Providing preservice teachers with an authentic forum to discuss issues which are relevant to the field of education will create knowledgeable and informed professionals who will become agents of change in our school systems in the United States.

References

Aleman, S. R. (1993). *Immigration education: A fact sheet.* (Report No. CRS-93-136-EPW). Washington D. C.: Congressional Library of Congress (ERIC Reproduction Service No. ED366694)

Archbald, D. (1994). *On the design and purposes of state curriculum guides: A comparison of mathematics and social studies guides from four state.* Report No. CPRE-RR-029). Brunswick, NJ: Consortium for policy Research in Education. (ERIC Reproduction Service No. ED374571)

Bercik, J. T. (1992). *Guiding future teachers into the twenty-first century.* Northeastern Illinois University. (ERIC Reproduction Service No. ED351284)

Borich, G. (1995). Becoming a teacher: An inquiring dialogue for the beginning teacher. *The Falmer Press Teachers' Library*, 7, 131-135.

Brott, P. & Kajs, L. (2000). *Developing the professional identity of first-year teachers through a "working alliance."* NAAC Homepage [On-line] Available:http://www.alt-teachercert.org/Working%20Alliance.html

Campus Update (1999). *Who are FAU students, 33*(4).

Cole, K. M. (1995, November). *Novice teacher efficacy and field placements.* Biloxi, MS: Annual Meeting of the Mid-South Education Research Association. (ERIC Reproduction Service No. ED393849)

Cordeiro, P. A. & Smith-Sloan, E. (1995). *Apprenticeships for administrative interns: Learning to talk like a principal.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Crawford, J. (1997). *Best evidence: Research foundations of the bilingual education act. NCBE Report.* (Contract No. T295005001). Washington D.C.: Office of Bilingual Education. (ERIC Reproduction Service No. ED408858)

Denton, D. R. (1993). *Minority medical school enrollment in the SREB States.* Atlanta, GA: Southern Regional Education Board. (ERIC Reproduction Service No. ED363193)

Firestone, W. A. (1990). Implementation, effects of state education reform in the 80s. *NASSP Bulletin, 74*(523), 75-84.

Furlong, J. (1997). Mentoring and developing practice in primary schools: Supporting student teacher learning in school. *Journal of Education and Teaching, 23*, 99-101.

Hall, R. H. (1968). Professionalization and bureaucratization. *American Sociological Review, 33*, 92-104.

Hunt, K. (1965). *Grammatical structures written at three levels.* Champaign, IL: National Council of Teachers of English.

Kerr, S., VonGlinow, M. A., & Schriesheim, J. (1977). Issues in the study of "professionals" in organizations: The case of scientists and engineers. *Organizational Behavior and Human Performance, 18*, 229-345.

Kleinsasser, A. (1988). *Language role call: 24 students present...23 are seventh graders.* Jackson, WY: Annual Meeting of the Northern Rocky Mountain Educational Research Association. (ERIC Reproduction Service No. ED299242)

Laitsch, D. (1998). *State evaluation of teacher preservice programs. Texas and Florida. An issue paper.* Washington D. C. : AACTE Publications. (ERIC Reproduction Service No. ED417192)

LaVelle, M. (1996). *The importance of learning English: A national survey of Hispanic parents.* Washington D. C. : Center for Equal Opportunity. (ERIC Reproduction Service No. ED405726)

Leggett, D and Hoyle, S. (1987). Peer coaching: One district's experience in using teachers as staff developers. *Journal of Staff Development, 8*(1).

McAllister, E.A. and Neubert, G.A. (1995). *Helping new teachers: Preservice peer coaching.* Washington D. C. : Office of Educational Research and Improvement. (ERIC Reproduction Service No. ED379720)

Miles, M. & Huberman, M. (1994). *Qualitative data analysis.* Thousand Oaks, CA: Sage.

Nieto, S. (1996). *Affirming diversity: The sociopolitical context of multicultural education.* (2nd. Edition). Longman: New York.

Ngyuen, T. (1987). *A summary of participation and achievement information as reported by state migrant education programs for fiscal year 1985. Vol. 1: Participation.* Department of Education, Office of Planning, Budget and Evaluation: Washington D. C. (ERIC Reproduction Service No. ED321930)

Ogle, D. (1986). K-W-L: A teaching model that develops active reading of expository text. *The Reading Teacher, 39*, 564-570.

Richards, J. (1992). *Beginning professionals' metaphors in an early field placement.* San Francisco, CA: Paper presented at the Annual Meeting of the American Educational Research Association. (ERIC Reproduction Service No. ED 348353)

Riley, R. W. (September, 15 1998). US Secretary of Education. *The challenge for America: A high quality teacher in every classroom.* Washington D. C.: Annual Back to School Address National Press Conference.

Routman, R. (1991). *Invitations: Changing as teachers and learners K-12.* Heinemann: Portsmouth, NH.

Saulpaul, C. and Badiner, K. (1997). *The Hyperstudio story.* Online: www.hyperstudio.com.

Sietsma, J. and Bose, J. (1995). *Characteristics of the 100 largest public elementary and secondary school districts in the United States: 1992-93. Statistical Analysis Report.* Washington D. C.: National Center for Educational Statistics. (ERIC Reproduction Service No. ED392850)

Snow, C. E. , Burns, S. M. and Griffin, P. (1998). *Preventing reading difficulties in young children.* Sponsored by Committee on the Prevention of Reading Difficulties in Young Children, National Reading Council

Sonneborn, R. (1970). *Friday night is papa night.* New York, NY: Viking.

Spatig, L. (1994). *Student teaching as social reproduction: An Appalachian ethnography.* New Orleans, LA: Paper presented at the Annual Meeting of the American Education Research Association. (ERIC Reproduction Service No. ED380408)

Trueba, E. (1998). Mexican immigrants from El Rincon: A case study of resilience and empowerment. *TESOL Journal, 7*(3), p.12-17.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 2, 21-32

East Baton Rouge, School Desegregation, and White Flight

Stephen J. Caldas
University of Louisiana

Carl L. Bankston III
Tulane University

Using archival, interview, and demographic data, the authors do an extensive historical analysis of school desegregation–and its possible consequences–in East Baton Rouge Parish, Louisiana. Enrollment trends of Black and White students are graphed and analyzed over the 32 year period from 1965 through 1997. During the period of "freedom of choice" desegregation from 1965 through 1980, the proportion of students in the parish's nonpublic schools actually decreased. Following court-ordered busing in 1981, however, there was massive White flight both to the parish's nonpublic schools and to two adjoining suburban school districts. The school system quickly went from majority White to majority Black. Recent costly efforts to reverse White flight have not worked in a system which was almost 70% African American in fall 2000. First-person accounts with affected stakeholders are provided.

Baton Rouge, the central city of East Baton Rouge Parish (EBR) and its associated public school district, is the second largest urban area in Louisiana. Like most school systems in the South, it was de jure segregated until 1954, after which time it went through several phases of desegregation. During the district's "freedom of choice phase," which lasted roughly from about 1965 to 1980, the trend was for White students to actually leave the district's nonpublic schools to attend its public schools. However, with the advent of more coercive desegregation measures in 1981, this trend abruptly ended and reversed itself. The system rapidly went from having a majority White to having a majority Black student population.

This paper analyzes fluctuating student populations over a 32-year period in terms of changing racial desegregation policies and practices. Based on our calculations, if the current rate of student White flight continues, East Baton Rouge parish will be an all Black school system by 2024.

White Flight

Our interviews with parents all over Louisiana indicate that most choose what they believe to be the best educational opportunities they can provide for their children. If they believe that sending their children from relatively high achieving schools to relatively low achieving schools could have negative academic consequences for their children, they usually opt for a more favorable educational alternative, if they are able. Our research indicates that majority African American schools have much lower achievement levels than majority White schools and that these schools tend to be disproportionately populated with students who are both economically disadvantaged and from single-parent families. These circumstances are highly correlated with lower student achievement for *all* students within schools (Bankston & Caldas, 1998a, 1998b; Caldas & Bankston, 1998, 1999). This–and not necessarily White racism–explains in part, we believe, the burgeoning nonpublic school population in East Baton Rouge parish, as well as some of the phenomenal growth in her two fast growing suburban parishes.

In previous research, we showed how the desegregation experience in Louisiana's largest public school system, New Orleans, evolved over time, and how White flight ultimately transformed the system into an almost all-African American school system (Caldas & Bankston, 1999a). In many respects, the situation of Baton Rouge is quite different from that of New Orleans and more akin to the rest of the South. First, Baton Rouge is not part of Catholic Louisiana (Caldas, 1992) and so does not share the strong historical tradition of parochial schools of its Latin counterpart to the south. Second, Baton Rouge did not follow the New Orleans pattern of gradual, piecemeal desegregation to avoid judicial mandates. As we show in this paper, by contrast with New Orleans, Baton Rouge took an approach to school desegregation that involved

---

Stephen J. Caldas is a professor in Educational Foundations and Leadership at the University of Louisiana in Lafayette. Carl L. Bankston, III, is an associate professor of sociology at Tulane University. Correspondence regarding this article should be addressed to Stephen J. Caldas, Dept. of Educational Foundations and Leadership, P.O. Box 43091, University of Louisiana, Lafayette, LA 70504 or by e-mail at caldas@louisiana.edu.

delay, followed by a drastic effort at coercion. Though there are experts who disagree (e.g., Orfield & Eaton, 1996), other important experts in school desegregation believe that coercive desegregation measures such as busing are counter-productive (Armor, 1995; Coleman, 1975; Rossell & Armor, 1996).

EBR Parish Sociodemographics

Though undergoing a demographic transition, Baton Rouge still had a majority White population through 1998 (U.S. Census Bureau, 1999b). The 1998 U.S. Census Bureau estimated that the population of Baton Rouge was 395,673 (up 4% from 1990) with 232,859 Whites (59%; down from 63% in 1990)) and 147,565 Blacks (37%; up from 35% in 1990). In the most recent complete census data (U.S. Census Bureau, 1990) in Baton Rouge, there were also large racial differences in income level. Among White households in East Baton Rouge, 20% had incomes below $15,000 per year. Half (slightly over 49%) of the Black households had yearly incomes below $15,000. Moreover, over one of every five Black households (22%) in Baton Rouge had yearly incomes below $5,000. These incomes translated into vastly different economic situations for White and Black schoolchildren. Only 7% of White children aged 5 to 17 lived below the poverty level in Baton Rouge. However, almost half (46%) of the Black school-aged children lived in poverty. Over one third of Black adults (35%) had not finished high school, and 14% of Black adults in East Baton Rouge Parish had completed less than ninth grade. Only 12% of White adults had not finished high school, and only 4% had not finished the ninth grade.

These statistics reflect the effects of the long history of racial oppression, some of which have been linked to the inferior education of African Americans under the "Separate But Equal" doctrine (Coleman et al., 1966). This history has left the much of the Black population of Baton Rouge, like much of the rest of the South, with high rates of poverty, single-parent families, and limited family educational backgrounds. The effort to overcome the oppressive past and to rectify a present in which racial discrimination has become structurally embedded has lent moral force to the struggle to desegregate the schools of Louisiana's capital. At the same time, though, Black social and economic disadvantages may have undermined this struggle. Our previous research has shown that avoiding schools in which poor children from single parent families are a dominant presence has meant avoiding schools in which there are large numbers of minority children (Bankston & Caldas, 1997, 1998a, 1998b; Caldas & Bankston, 1998, 1999b)–a factor which might be keeping middle and upper income Whites and Blacks from placing their children in predominately African American schools.

The Desegregation Struggle in Baton Rouge

The constitutionality of Baton Rouge's de jure segregated school system was first challenged in 1956 in the case of *Davis et al. v. East Baton Rouge Parish School Board*, litigation which has been active in the federal court system ever since (1960/1961/1961/1967/1979/1983/1996), making it the longest unsettled desegregation suit in U.S. history. The "Freedom of Choice" approach to desegregation was subsequently adopted in the 1960s in East Baton Rouge, ending de jure segregation and allowing Black and White students the freedom to attend each others' schools. However, "freedom of choice" rarely resulted in significant desegregation of either schools or school systems. (For an in-depth case study highlighting problems of "freedom of choice," see Causey, 1999).

According to plaintiffs in the original desegregation lawsuit (which included the NAACP), Baton Rouge schools continued to be segregated on a de facto basis throughout the 1970s (Baird & Luster, 1990). In 1981, Federal District Court Judge John Parker was convinced by the plaintiff's arguments and found that the East Baton Rouge Parish School Board had maintained what he termed a "dual school system" for 20 years. He subsequently designed an extensive desegregation plan for Baton Rouge schools. Importantly, however, the Justice Department strenuously opposed Parker's plan from the beginning. The Department argued that it was causing Whites to flee from the system (Thornton, 1982). The White exodus that ensued probably even caught the Department of Justice by surprise.

In the face of much opposition, Parker ordered 15 East Baton Rouge schools closed in May 1981 in an effort to achieve racial balance. Under his plan, formerly White and formerly Black schools which remained open were to be paired or clustered, and students bused to schools in their "cluster" based on the need to create racial balance (Baird & Luster, 1990). Parker's desegregation orders provoked massive resistance and an immediate exodus of White students from the public school system. The president of the Central Middle School Parent-Teacher Organization, whose daughter would have been transferred to the (largely Black) Scottlandville school under the desegregation plan proclaimed at the time that, "She will not do that. Private schools are starting up every day"("Parents Enraged," 1981). Indeed, by the end of the first year of Judge Parker's plan, private schools in East Baton Rouge Parish were noting that their waiting lists were long and growing longer, so long, in fact, that new schools sprouted up almost overnight to accommodate the sudden demand for nonpublic education. One of the city's largest kindergarten through twelfth-grade nonpublic schools, Parkview Baptist, dates its founding to that first year of forced busing in 1981, and by 1998 it enrolled 800 students in its high school alone. Enrollment figures in nonpublic parish schools jumped by approximately 2000 in that first year of busing (Louisiana Department of Education, 1965-1997). Moreover, according to subsequent enrollment trends, those students never returned. Using data released annually by the Louisiana Department of Education, we now turn to an examination of the student demographic trends in East Baton Rouge from1965-1997.

Baton Rouge: What Happened?

In order to answer the question of what actually happened during East Baton Rouge's desegregation efforts, we track and present longitudinal data on the parish and its metro area for three variables: the percent of all parish public school students who are African American, the percent of all White students enrolled in Baton Rouge nonpublic schools, and the percent of all White students in the entire metropolitan area who were enrolled in public schools in parishes immediately *outside* of Baton Rouge.

For the purposes of this study, we define the Rouge Parish metropolitan area as including East Baton Rouge parish, and the two adjacent suburban parishes of Ascension and Livingston. An examination of the population trends quickly reveals that most of the growth in Baton Rouge's metropolitan area is taking place in these two parishes. Indeed, in the mid-1990s, Ascension and Livingston were among the three fastest growing parishes in Louisiana. (The other was the New Orleans metro "White flight" parish of St. Tammany). In 1998, there were an estimated 71,628 residents in Ascension Parish, an increase of 23% over 1990. Twenty-five percent of Ascension Parish was Black in 1998 (a decrease of 35% from 1990). Livingston Parish had an estimated population of 88,104 in 1998, also a 25% increase over its 1990 population. Six percent of Livingston's residents were Black in 1998, a decrease of fully 41% from 1990 (U.S. Census Bureau, 1999b).

Thus, we see fairly stark contrasts between the demographics of East Baton Rouge parish and her two fast-growing bedroom communities. While the population of Baton Rouge grew only slightly from 1990 to 1998, the percentage of all Black residents increased sharply, up 11%. However, while both Ascension and Livingston parishes grew rapidly during the same 8-year period, the proportion of African Americans in both parishes dropped sharply.

Focusing on school demographic changes, we graph the fluctuation of the racial composition of Baton Rouge public and nonpublic schools and the racial composition of the "White flight" parishes over a 32 year period. As can be seen in Figure 1 from the line representing the percent of all East Baton Rouge parish public school students who are Black, a minority (39%) of those in this public school system were Black in 1965.

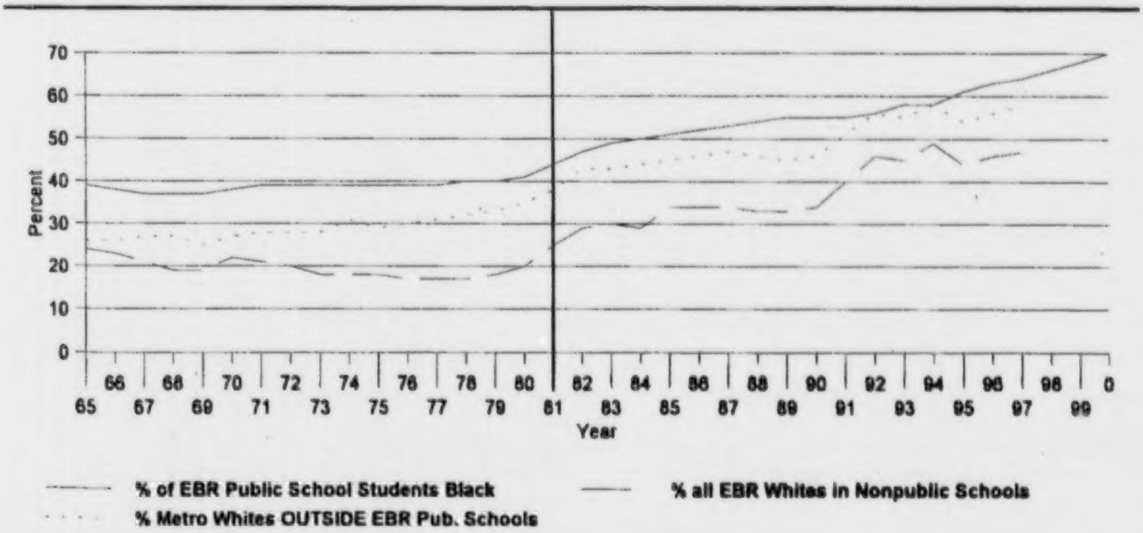


Figure 1: Exodus of White Public School Students from EBR Parish, 1965-2000.

Only a small percentage of Baton Rouge's White students were in nonpublic schools in 1965 (24%). Also, the vast majority of the White students in the three-parish Baton Rouge metropolitan area were specifically enrolled in EBR Parish schools (74%).

Following the initial law suit to force the desegregation of Baton Rouge's schools, the school board allowed schools to desegregate voluntarily during the 1960s, employing the "freedom of choice" approach to desegregation. However, "freedom of choice" as a desegregation tool was ruled unconstitutional in the 1969 U.S. Supreme Court case *U.S. v. Greenwood Municipal School District*. Since Baton Rouge's neighborhoods were (and are) largely segregated by race, with most Blacks living in the northern part of the city and most Whites living in South Baton Rouge, the city's public schools continued to remain largely segregated by race until the court-order of 1981. Having this extended period of voluntary desegregation, when children in the city attended schools located in their own neighborhoods but were not prohibited by race from attending any public school, provides a unique research opportunity. It allows us to observe trends within the system during the period of "voluntary" desegregation following the end of de jure segregation in Baton Rouge's schools, and then to compare these trends with the period following more forcible desegregation in 1981.

We can see in Figure 1 that from the mid-1960s through 1980, the percentage of Blacks in Baton Rouge public schools remained relatively constant. We calculated that the percentage of the system that was African American was increasing at an average of only 0.21 percentage points annually, or roughly only a fifth of a percent per year during this 15 year period (see Table 1).

Table 1
Annual Percent Growth in Proportion of African American Students in East Baton Rouge Parish Public School System

Range in Years	Average Annual Percent Change
1965-1980	.21 percent increase per year
1981-1990	1.1 percent increase per year
1991-1997	1.43 percent increase per year

Also, during the same period, the total percentage of White students in the metropolitan area outside of Baton Rouge increased only slightly. Significantly, within the district there was a noticeable trend for White students to move **out** of the nonpublic schools and **into** EBR's public schools. In Table 2 we calculated that East Baton Rouge parish nonpublic schools were actually **losing** enrollment at a rate of about 0.36 percentage points per year until 1980.

Table 2
Annual Percent Growth in Proportion of All White Students Enrolled in East Baton Rouge Parish in *Nonpublic* Schools

Range in Years	Average Annual Percent Change
1965-1980	.36 percent *decrease* per year
1981-1989	.96 percent increase per year
1990-1997	1.36 percent increase per year

Thus, we can surmise that in general, Whites were relatively content with the public school system during this time frame, even to the extent of Whites leaving their nonpublic schools. This is especially noteworthy given that this period coincides with Black students attending formerly all White schools for the first time in the twentieth century. By contrast, in some other areas of the country (e.g., New Orleans) the same conditions resulted in continuous White flight to segregated nonpublic schools. These figures for EBR suggest that most Whites were willing to attend schools with African Americans--as long as Black students were entering majority White schools in relatively small numbers. This condition, though, was the very reason that the court began to pursue a more aggressive desegregation policy.

With the onset of court-ordered, forced desegregation and busing in 1981, the educational situation in the Baton Rouge metro area began to change drastically. As seen in Figure 1, there was an immediate and precipitous flight of White students to Baton Rouge's nonpublic schools (note: vertical line represents 1981). The Department of Education data indicate that in the first year of court-ordered busing alone, the East Baton Rouge public school system lost 7,000 White students, or the equivalent of four large public high schools. The African American proportion of the student body jumped from 41% in 1980, to 44% in 1981. Since the parish nonpublic school population jumped almost 2,000 students in those same two years, we can be reasonably confident that the massive decrease in the number of students from the public schools was not due primarily to other demographic factors, such as the "baby bust." Indeed, from the very start, the "White flight" situation was perceived as so grave a peril to the health of the system that even the U.S. Justice Department, which had early misgivings about the desegregation plan anyway, requested that the court reconsider its earlier position (Thornton, 1982). The court pressed on.

By our calculations, the White population of East Baton Rouge's **nonpublic** school student population has increased by an average of 1.4 percentage points per year since forced desegregation of the public school system (see Table 2). In short, when one considers that prior to forced desegregation Whites actually seemed to favor the public schools, as is evident by the steady trickle of students from the nonpublic schools, this seems to be fairly reliable evidence of deep disenchantment with the decision to desegregate the schools through sudden and massive shifts of Black and White students.

Figure 1 shows that there was a brief period during the late 1980s when White student loss from Baton Rouge's public schools leveled off, before accelerating once again. According to a central office administrator of the East Baton Rouge public school system, the temporary stem in the exodus of White students was because of a short-lived experiment where the school board allowed "controlled choice." It was a strategy by the then new superintendent, Dr. Bernard Weiss, to reduce mandatory busing and "regain the trust and participation of families and local agencies who no longer supported the efforts of public education" (Baird, 1990). The school redesign "controlled choice" plan included setting up popular educational programs, like a dyslexia program, in predominantly minority schools, and allowing White children from other school districts within the parish to attend. These magnet programs were unavailable anywhere else. Thus, the disadvantages associated with an increasing minority presence in schools, such as more poor students and more students from single parent households, could be somewhat counterbalanced by providing special educational opportunities to attract White families. However, after a couple of years of modest success, the board was unable to expand the strategy to additional schools due to funding restrictions, as well as a lack of interest among some predominantly White schools that were not (according to our conversations with school officials) particularly happy about having Black students bused in (J. Baird, personal communication, December 18, 1998).

The rate at which the district was becoming a primarily Black district at the end of the twentieth century was accelerating. The rate jumped from an average of just .22 percentage points per year during the freedom of choice period, to an average of 1.1 percent per year from 1981 to 1989 during the first 8 years of coercive desegregation. It increased at an even faster rate during the first seven years of the 1990s, accelerating to an average of 1.4 percentage points per year. Indeed, from 1996 to 1997, the district lost 1000 White students but gained more than 250 Black students.

As can be seen in Figure 1, Whites were not just fleeing to the nonpublic school system but appeared to also be heading toward the suburban parishes of Ascension and Livingston. In the 1996-97 school year, fully 56% of all White students in the tri-district area were enrolled in these two fast growing metro parishes. Based on our calculations, the percentage of metropolitan area Whites who were in these parishes' suburban public schools, rather than in EBR schools, increased at an average rate of 1.2% per year from 1981 to 1997. This compares to an average rate of increase of only 0.56 percentage points per year prior to court-ordered desegregation.

Though there has obviously been out-migration of Whites from East Baton Rouge parish, this was not the primary cause for the initial, disproportionately rapid growth in the proportion of Blacks in the parish's public schools. According to estimates from U.S. Census data (1998), the proportion of Blacks in the parish increased from only about 29% in 1970, to 37 percent in 1996. However, during this same 28-year period, the increase in African American representation in the public schools was much greater, from 38% to 63%. In other words, the proportion of Blacks in Baton Rouge's public schools grew more than three times faster than the Black growth rate in the parish overall. The overall White population, too, actually increased in numbers, from 205,528 in 1970, to an estimated 241,271 (U.S. Census Bureau, 1998), though these numbers would have been arguably higher, perhaps much higher had the EBR school system been more attractive to Whites than were the two suburban systems.

With the total collapse of the "school redesign" plan in 1991, one can see from Figure 1 that White flight from Baton Rouge's public schools continued unabated through 1997 (by the upwardly rising line), with only one additional, slight reprieve in 1994-1995. There was a particularly precipitous drop in White student enrollment from 1995-96 to 1996-97, when the White student population declined by more than 1500—the equivalent of an entire large high school's student population.

During this same one-year time frame, the percentage of the school district that was Black increased from 61 to 63 percent. We calculated that on average, from 1980 to 1997, the percentage of African American students in the EBR system increased 0.95 percentage points a year, or just under one percent annually (see Table 2). This compared to only 0.22 percentage points for the years from 1965 through 1980, the year preceding Federal court-ordered desegregation.

As telling as they are, the numbers are only part of the story of the unintended consequences of the sudden, coercive turn in the desegregation of Baton Rouge schools. Our extensive interviews with individuals who experienced the desegregation of Baton Rouge's schools first hand can give deeper insight into how families responded to these events. We believe that their first-hand accounts are fairly representative of the general sentiment of many White parents and citizens toward EBR's desegregation experiment, though the sampling admittedly was not scientific. The researchers heard comments about how the shifting school zones, which sometimes sent children from a neighborhood school across from their house on an hour and a half or more bus ride to another school across town, hurt "community ownership" in schools.

In one representative example, a student football player who attended a city Catholic high school prior to court-ordered busing shared how much public-school community spirit he observed at a football game against a city public high school (Broadmoor High) in 1977. "When we broke through the banners in the end zone, the stands were filled and there was standing room only " (V. Lagattuta, personal communication, 1999). The former student continued that when he returned to watch a game between the same two teams in 1981, after the dismantling of neighborhood schools, "I was in shock . . . the fan base [at Broadmoor] had gone from standing room only to almost total abandonment. I did feel sorry for the kids on the Broadmoor football team."

A Red Cross volunteer at her children's elementary school in the early eighties said that there was an enormous difference after Black children were bused across town to the formerly predominately White school:

> I personally saw more kids come to the clinic, and often couldn't get in touch with a family member to get the sick children. It seems to me most of the ones coming in there were the students bused in from outside the neighborhood. Broken homes, one parent households, or living with grandparents who could not retrieve them. I often spoke with the teachers I knew and they all commented about the problems caused by the parents not being from the neighborhood. They just couldn't seem to be involved enough. The teachers couldn't get enough volunteers in the classroom to help for special functions. I also spoke to some of the Black mothers and they were not happy about having their children being bused out of their neighborhoods. It wasn't convenient anymore for them to attend school functions because of the distance and traffic. Very few people were happy with the situation" (D. Walden, personal communication, December 17, 1999).

In 1999, in the 90% White community of Kennelworth, the local elementary school had only seven Whites enrolled. "One can see private school Whites waiting outside every morning for their school buses," one local authority reported to us (C. Tolbert, personal communication, August 26, 1999). In the mostly White southeastern part of East Baton Rouge parish, one of the last battlegrounds of desegregation litigation was taking place at Wedgewood Elementary School in 1999, where the unsettled situation could be the reason for the school population going from 10% to 30% Black in a period of five years. Though for the moment the majority of the parish population is White, with almost 50% of all White students in nonpublic schools, sociologist and demographer Charles Tolbert characterized the situation as "very fluid" (C. Tolbert, personal communication, August 26, 1999). In a report for the school board, Tolbert (1999) projected that by the year 2005, 50% of the Baton Rouge school-aged population will be Black. "Baton Rouge appears to be going the way of New Orleans," Tolbert remarked (C. Tolbert, personal communication, August 26, 1999).

In an almost desperate attempt to stem the flow of White students from the Baton Rouge system, the 1996 desegregation plan did away with much (though not all) forced busing (many Black and White parents still complained that they could not send their children to the closest neighborhood school). It also created 24 separate magnet programs to attract Whites to majority Black, inner-city schools. Half of the programs failed to attract even 10 White students (King, July 14, 1999). It was not for lack of money that these magnet programs failed to work. Between 1996 and 1999 the school board poured $6.8 million dollars into special programs designed to appeal to White students. Nor was it for lack of creativity that certain programs did not work. For example, even though Louisiana is in the midst of a French language revival and some districts have long waiting lists for certain foreign language programs, a newly created foreign language magnet program in a majority Black Baton Rouge school did not attract even one White student (King, July 14, 1999). A program's ability to attract potentially wary parents and students appears to depend on more than money and course content. Although parents do consider benefits such as magnet programs, the location and the composition of student populations apparently also figure heavily in their calculations.

We see, then, that the trend toward a majority Black school district in Baton Rouge took place almost immediately following the implementation of court-ordered desegregation in 1981. The single best indicator that the White flight was a result of the sudden changes brought about by an aggressive desegregation effort and not by a tendency toward suburbanization, was the drastic shift of the White student population from public to nonpublic schools that began at this time. Whereas White student enrollment in Baton Rouge nonpublic schools had actually been decreasing prior to official desegregation, following Judge Parker's court order we see that this trend not only ceased, but actually accelerated markedly in the other direction. In 1965, 24% (11,375) of EBR Parishes' White student population was in nonpublic schools. During the next 15 years, when voluntary racial desegregation was the official school board policy, the percentage of White students in the city's nonpublic schools decreased to only 20% (9913). In the very next year, 1981, when court-ordered busing went into effect, the percentage of Whites in Baton Rouge's nonpublic schools jumped to 25% (to 11,837). By 1997, this percentage had climbed to 47% (17,283) at the same time that the White EBR population was actually in decline. This strongly suggests that the primary cause of this enormous shift of White students from the public to nonpublic schools was a direct result of the dismantling of neighborhood schools and the sudden changes that followed coercive desegregation.

As impressive as the White flight to nonpublic schools was, there appeared to be an equally impressive migration of Whites to Baton Rouge's suburban sister parishes during this time frame as well. And while it is harder to link this population shift to Baton Rouge's judicially mandated effort at desegregation, the coincidence in timing strongly suggests that a link does exist, and that the link may be a strong one indeed. Baton Rouge's bedroom communities of Ascension and Livingston Parishes—two of the three fastest growing parishes in Louisiana—went from containing only 26% of all White students in the tri-parish metro area in 1965, to enrolling fully 56% of the metro area's White student population by 1996. This was during the same time that the overall population of East Baton Rouge was growing rapidly.

A comment by the principal of Denham Springs Junior High, in Livingston Parish's largest city, seems to characterize well the local population's perception of the extraordinary growth in his parish. In an interview with the longtime Livingston educator, he commented to one of the authors that Livingston's growth, ". . . is almost exclusively driven by White flight and the initial location of new hires [workers] for industry in EBR who will not live where they work" (W. Smith, personal communication, January 27, 2000). He believes that avoidance of Baton Rouge's schools is the primary reason for people moving into his parish.

All these statistics have not been lost on an often divided school board which has been struggling to win back the support of White residents. After the school redesign plan of the late 1980s and early 1990s fizzled, the board proposed an ambitious plan to rebuild many of the city's crumbling schools in addition to building several new schools. It was heralded as the most expensive public school expenditure plan in the history of Louisiana and indeed the entire U.S. The board championed the plan as the last chance to salvage a disintegrating system. In order to fund the $2.2 billion dollar program, the board went to the city's voters with a tax and bond plan. Baton Rouge voters defeated the proposal by a margin of two to one. In the city of Baker, in the northern part of the parish, voters were so dissatisfied with the school district that they voted to create a new district because these rural residents "objected to having their children bused up to 30 miles to school" (Anderson, 1995, p. A-15). Another constitutional amendment was passed by the voters of Louisiana in October 1999 granting another small city in East Baton Rouge parish, Zachary, the right to create its own school system.

There is obviously some support for the Baton Rouge public school system, as a much more modest tax hike to improve parish-wide education did pass in 1998. However, based simply on the massive, continuous out-flow of White students into the parish's nonpublic schools and into public schools of the surrounding parishes, it seems that confidence in the school system among Whites is steadily eroding. In a 1999 draft of a report on desegregation in Baton Rouge nationally recognized desegregation expert Christine H. Rossell remarked: "I do not believe I have ever been in a school system where the schools were in such poor condition as a result of taxpayer non-support" (Rossell, 1999, p.6). Recalling the economic statistics cited above and the racial inequality they highlight, White taxpayers are the primary potential source of local funding for schools. Future appeals to the voters of East Baton Rouge parish for public school funding are increasingly likely to fall on deaf and uninterested ears, especially the ears of those who have opted to pay the higher costs of nonpublic education and who would resent double payment.

Continuing De Facto Segregation and the Decline of a School System

As already mentioned, a court-approved consent degree in 1996 ended much forced busing in East Baton Rouge. Between 1996 and 1999 the school board spent $27 million on desegregation efforts, including funding for special accounts, called "equity accounts," for historically Black schools. In fact, some majority Black schools were receiving so much funding, including special monies for technology and traditional Title 1 funding for having many students from poor families, that the superintendent was quoted as saying, "The principals are telling me they're finding it difficult to decide what else they need beyond what they've already bought" (cited in King, July 13, 1999). Meanwhile, none of these financial incentives has stopped the rate of White flight or the increasingly bitter quarreling among school board members, the local NAACP leadership, and the federal judge who initiated the desegregation process. Between 1996, with the new consent degree designed to end White flight and the opening of schools in 1999, the East Baton Rouge school system went from 63% to 65% Black (King, July 15, 1999). By 1999, not only White enrollment, but Black enrollment as well was significantly lower. On the opening day of school in 1999, the school system registered the smallest number of students in more than two decades. A total of 2300 fewer students showed up for the first day of school in 1999 than on the first day of school in 1998 (King, August 20, 1999).

Even as the system was literally "disintegrating," the heated rhetoric on both sides seemed to be taking on a nonsensical quality. The local superintendent of education, Dr. Gary Mathews, maintained that the system had demographically arrived at a point where further racial desegregation was almost physically impossible. This contention was supported by Tolbert (1999) who projected that if rates of White flight continued, Baton Rouge public schools would be totally African American by 2020. Tolbert said that with virtually each court consent decree or new desegregation plan, the exodus of Whites from the system has increased, as Whites react to uncertainty in the system by simply leaving it (C. Tolbert, personal communication, August 26, 1999).

Superintendent Mathews has pleaded for an end to the longest desegregation lawsuit in history, so that the system would be free to save itself without outside interference. Even Louisiana Governor Foster indicated he wanted to "pay a personal call"on Judge Parker to discuss ending federal involvement in the EBR system (Redman, August 16, 2000). Many Black parents agree with him. Larry Galloway, an African American with two children in Baton Rouge public schools, has teamed up with 25 other parents to end the more than four decades old desegregation suit. He observed, "I think we're at a point where Baton Rouge probably cannot be fully desegregated" (cited in King, March 25, 1999, p. 1A). However, the local NAACP leader strenuously took issue with any suggestion that the courts back out, contending, "There is no way we're going to agree to removing judicial involvement in this system as long as we don't feel our children can get a fair shake without it" (quoted on WBRZ-TV 10 p.m. news, August 19, 1999).

At the end of 1999, the Baton Rouge school board voted to meet once again with Judge Parker, where they wanted to propose such measures as reinstating certain magnet/gifted programs targeted specifically at Whites and then offering to bus White children directly from their homes to these programs (King, August 29, 1999). Ironically, similar programs were dismantled only a few years earlier because they created concentrations of Whites. Meanwhile, White and, increasingly, Black parents silently continue to remove their children from a system taking on the inner-city characteristics of William Julius Wilson's "truly disadvantaged" (Wilson, 1987).

One parent who did not remove her children from Baton Rouge's public schools as a matter of principle —but now wishes she had—is the former head of Louisiana's influential Public Affairs Research Council. Jackie Ducote was an ardent education reformer and public school supporter who felt that she could not go before the state legislature and lobby for public schools if she herself did not set an example by keeping her children in Baton Rouge's public schools (McClain, 1999). Her children endured the upheaval of the early days of forced busing, and Ms. Ducote now admits they were not progressing as well as she would have liked, and her sons eventually found themselves in remedial college classes. They never obtained college diplomas. She confesses that "I probably sacrificed Chip and Drew [her sons] for my principles . . ." (McClain, 1999, p. 17A).

The same sentiment was expressed by the earlier quoted Red Cross volunteer who initially left her son in the public schools because "we didn't feel we could afford it [private schools] at the time" (D. Walden, personal communication, December 17, 1999). However, when she obtained a paying job several years later, she removed her son from a city high school where "he was miserable and became withdrawn" and enrolled him in a private school. She commented that "he loved it . . . I wished I had taken him out of the public school system sooner. I very strongly feel that the deseg [sic] in Baton Rouge was a failed experiment" (D. Walden, personal communication, December 17, 1999).

Consequences of Changing Racial Compositions

Why would changing racial compositions of Baton Rouge schools necessarily equate with inferior education? Or, as East Baton Rouge school board member Patrice Niquille is quoted as saying during a board meeting, "I'm concerned because I'm getting the message here tonight that too many Black students in our (school system) is bad . . . and I don't think we want to send that message to our community" (cited in King, July 13, 1999). It is a valid question, and, as noted above, our research suggests that the answer is in part due to the strong association between race, socioeconomic status, family composition, and lower levels of achievement (Bankston & Caldas, 1996, 1997, 1998a, 1998b; Caldas & Bankston, 1997, 1998a, 1998b, 1999b)—not necessarily due to White racism, although racism undoubtedly continues to be part of the problem.

Simply put, race, class, and education are closely inter-related in Baton Rouge, as they are in much of American society. In 1990, whereas only nine percent of all Baton Rouge Whites lived in poverty, the rate was 38% for Blacks, or more than three times higher according to 1990 U.S. Census figures. As we pointed out above, there were significant racial gaps in educational attainment as well. Results from Louisiana's Graduate Exit Examination show clear racial differences in achievement levels. From data tapes provided by the Louisiana Department of Education (1990), we calculated that in 1990—a mid-point in the desegregation experiment— East Baton Rouge Whites did significantly better than Blacks on all three tenth-grade components of the test. This Black-White gap continued throughout the decade. Table 3 presents Black-White gap in East Baton Rouge schools in 1990, 1994, and 1999.

Table 3
Average Percent Correct on Tenth Grade Components of 1990, 1994, and 1999 Louisiana Graduation Exit Examination: East Baton Rouge Parish

	Math	Language Arts	Written Composition
1990			
Blacks (N=1,822)	59.%	69.4%	74.9%
Whites (N=1,729)	75%	81.9%	85.9%
1994			
Blacks (N=2,425)	54.6%	67.8%	72.1%
Whites (N=1,807)	73.9%	82.5%	83.1%
1999			
Blacks (N=2,566)	54.6%	73.0%	79.9%
Whites (N=1,611)	72.0%	85.5%	86.8%

Much of these data were made available to the general public. Indeed, beginning in 1990, the Louisiana Department of Education provided every public school parent in Louisiana with a copy of their child's school "report card." In other words, parents are well informed regarding the academic achievement levels of students in Baton Rouge's schools.

The results of the 1999 LEAP test, which were used to actually rank schools into five categories according to academic performance, were also well-publicized. There were no EBR schools in the highest tier, and only one school in the second tier. It was a middle school, located in an upscale suburb of the school district. It also had the second smallest minority population of any school in the district, with only 19% Black students.

As we move down the scale of performance among Baton Rouge schools in 1999, the minority populations and the low-income populations of schools grow steadily larger. Most EBR schools were classified as either "Academically Below Average," or "Unacceptable," the two lowest of the five tiers. These schools were made up overwhelmingly of minority students and low-income students.

It begins to become evident why, apart from the inconvenience of busing, White parents in East Baton Rouge would be upset by suggestions of a massive reshuffling of students. There is a great deal of variation in performance levels among schools. Moreover, this variation is closely tied to the racial and socioeconomic make-up of schools. Clearly, parents who are concerned with placing their children in the most advantageous educational circumstances will want their children in schools with few low-income students and few Black students. These are the schools the children of White middle-class families would normally attend without difficulty, unless there is an official mandate to redistribute students. From 1981 onward, the mothers and fathers of middle-class White children were continually faced with the possibility that their children would be taken out of "good" neighborhood schools and moved to low-performing schools.

What are the advantages of settling outside of East Baton Rouge, in the areas generally identified as "White flight" areas? Although a majority of schools in Baton Rouge are classified as "Below Average" (based on test scores) by the Louisiana State Department of Education, 93% of the schools in the surrounding suburbs are classified as "Above Average" or better. The schools in the White flight areas do appear to have a fairly high degree of racial segregation: Among the highest-classed schools, according to the Louisiana Department of Education, all had very small percentages of minority students. The "Below-Average" schools in the suburban fringe all had student bodies that were almost entirely Black. From the point of view of egalitarian ideals of social justice, this is very disturbing. But from the point of view of a parent with a school-age child seeking the best academic environment, the implications for action are both obvious and inconsistent with egalitarianism. Inside the city and outside, the best schools tend to be those with the smallest minority populations. There is a strong correlation among the percentage of Black students, percentage of low income students, and low school scores in EBR. In the Baton Rouge area, the correlation (*r*) between percent minority and percent poor is .78, the correlation between percent minority and school test scores is -.76, and the correlation between percent poor and school test scores is -.72 (not shown in a table). In other words, the more minority students there are in a school, the more low-income students there are in that school. The more low-income and minority students in a school, the lower the school's measured performance.

Our interviews with parents all over South Louisiana have convinced us that most parents choose what they believe to be the best educational opportunities they can provide for their children. If they believe that sending their children from relatively high-achieving schools to relatively low-achieving schools could have negative academic consequences for their children, they will likely opt for a more favorable educational alternative, if they are able to provide one. This explains, we believe, the burgeoning nonpublic school population in EBR parish, as well as some of the phenomenal growth in her two fastest growing suburban parishes.

What can be done to stem the flow of White—and arguably middle class Black—students from the EBR public school system? Our research leads us to believe that desegregation seems to work best where the minority population is in the minority. As minority figures increase beyond a certain percentage, the majority White population—whether rightly or wrongly—begins to leave the public schools. Perhaps we can learn much from Baton Rouge. We see that after 1965, when schools could not refuse African Americans, yet prior to 1981, when parents were allowed to send their children to their neighborhood public schools, the system's public schools flourished. Nonpublic school enrollment decreased. However, the sudden death of neighborhood schools in 1981 marked the slow deterioration of Baton Rouge's school system. Had the federal courts never ordered forced desegregation, it seems highly likely that the system would still be a healthy one, though it would probably also still have racially identifiable schools. But from a practical perspective, this would still be better than the current plight of the system: it still has racially identifiable schools, but they are now in a crumbling, financially strapped system with virtually no support from the segment of the community with the most resources. In stark contrast to Baton Rouge, the school board of Lafayette Parish, Louisiana opted to keep its neighborhood schools intact during the flurry of desegregation activity in the 1970s—an action which may have stemmed massive White flight (Caldas, 1999).

Though, regrettably, it may be too late to make Baton Rouge's schools once again attractive to large numbers of middle and upper-middle income families—regardless of their race—we wonder if perhaps a return to the neighborhood school concept is a compromise well worth the trade-off of having some racially identifiable schools—which Baton Rouge would have regardless—for support from the socioeconomically advantaged segment of the community. Is it possible that this compromise may be the best chance we have of salvaging viable urban public education in many parts of the United States—including Baton Rouge?

References

Alexander, K. (1980). *School law*. St. Paul: West Publishers.

Anderson, E. (1995, October 25). Separate school districts may draw court challenge. *The Times-Picayune*, A15.

Armor, D. J. (1995). *Forced justice: School desegregation and the law*. New York: Oxford University Press.

Baird, J. (1990). *Examining parent selection in the East Baton Rouge controlled choice plan*. Unpublished manuscript, Louisiana State University.

Baird, J., & Luster, J. N. (1990). *Initiating school and district level change through restructuring in East Baton Rouge Parish*. Unpublished manuscript, Louisiana State University.

Bankston, C. L., & Caldas, S. J. (1998a). Family structure, schoolmates, and the racial inequalities in school achievement. *Journal of Marriage and the Family, 60*(3), 715-723.

Bankston, C. L, & Caldas, S. J. (1998b). Family structure, schoolmates, and the racial inequalities of Schools. *Sociological Spectrum, 18*, 55-76.

Bankston, S. J., & Caldas, S. J. (1997). The American school dilemma: Race and scholastic performance. *Sociological Quarterly, 38*, 423-429.

Bankston, C., & Caldas, S. J. (1996). Majority Black schools and the perpetuation of social injustice: The influence of de facto segregation on academic achievement. *Social Forces, 75*(2), 535-555.

Caldas, S. J. (1999, November). *A comparison and contrast of the desegregation experiences of three Louisiana school districts: The City of Monroe, Ouachita Parish, and Lafayette Parish*. Paper presented at the 1999 annual meeting of the Mid-South Educational Association, Point Clear, AL.

Caldas, S. J. (1992). The politics of welfare reform, religion, education & racism: Evidence from three Southern elections. *Southern Studies, 3*, 1-14.

Caldas, S. J. & Bankston, C. (1998a). Female-headed families and school environments: Does the family structure of schoolmates make a difference in academic achievement? *Louisiana Educational Research Journal, 23*(1), 101-128.

Caldas, S. J. & Bankston, C. L. (1998b). The Inequality of separation: Racial composition of schools and academic achievement. *Educational Administration Quarterly, 34*(4), 533-557.

Caldas, S. J., & Bankston, C. L. (1997). The effect of school population socioeconomic status on individual student academic achievement. *Journal of Educational Research, 90*, 269-277.

Caldas, S. J. & Bankston, C. L. (1999a, March). *The promise and failure of school desegregation in Baton Rouge and New Orleans*. Paper presented at the 1999 annual meeting of the Louisiana Educational Research Association, New Orleans, LA.

Caldas, S. J., & Bankston, C. L. (1999b). A multilevel examination of student, school, and district-level effects on academic achievement. *Journal of Educational Research, 93*(4), 91-100.

Causey, V. E. (1999). 'Drafted into the front lines': Teacher efficacy during school desegregation in Columbus, Georgia, 1968-1975. *Research in the Schools, 6*(2), 9-24.

Coleman, J. S. (1975). Racial segregation in the schools: New research with new policy implications. *Phi Delta Kappan, 57*, 75-78.

Coleman, J. S. Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of Educational Opportunity*. Washington, DC: U. S. Government Printing Office.

Davis et al. v. East Baton Rouge Parish School Board. 721 F.2d 1425 (5th Cir. 1983).

Johnson, D. W., Picard, P. R., & Quinn, B. (1981). *Churches and church membership in the U.S.* Washington, DC: Glenmary Research Center.

King, K. (1999, March 25). Parents target desegregation suit as obstacle. *The Advocate*, p. 1A, 6A.

King, K. (1999, April 13). EBR Hopes to benefit from teacher pay hike. *The Advocate*, 7B.

King, K. (1999, July 13). Expert: BR losing its White students. *The Advocate*, 1A, 4A.

King, K. (1999, July 14). Half of EBR magnets failed to attract 10 White students. *The Advocate,* 1A, 4A.

King, K. (1999, July 15). New desegregation concept could lead to legal challenges. *The Advocate*, 1A.

King, K. (1999, August 20). EBR awaits holiday count. *The Advocate*, 1A, 14A.

King, K. (1999, August 29). Board hopes to entice White students to other schools. *The Advocate*, 1A, 5A.

Louisiana State Department of Education. (1990). *Machine Readable Data Tape of Test Results of Louisiana Educational Assessment of Progress*. Baton Rouge, LA.

Louisiana State Department of Education. (1965-1997). *Annual Financial and Statistical Report*. Baton Rouge: Louisiana Department of Education.

McClain, R. (1999, April 25). Ex-PAR chief doesn't believe education can reform itself. *Sunday Advocate*, 1A, 16A, 17A.

Orfield, G., & Eaton, S. E. (1996). *Dismantling desegregation: The quiet reversal of Brown v. Board of Education.* New York: The New Press.

Parents enraged by school mix order. (1981, May 5). *The Times-Picayune,* section 5, p. 7.

Parkview Baptist Church. (No date). [On-line] Available: http://www.parkviewbaptist.com/hs01.html

Redman, C. (2000, August 16). Foster considers talk with judge in school suit. *Advocate Online* [Online], Available: http://www.theadvocate.com/

Rossell, C. H. (1999). *Improving the voluntary desegregation plan in the Baton Rouge school system.* A report to the court in the case of *Davis, et al. v. East Baton Rouge Parish School Board et al.*

Rossell, C. H., & Armor, D. J. (1996). The effectiveness of school desegregation plans, 1968-1991. *American Politics Quarterly, 24*, 267-302.

Shinkle, P. (1996, July 29). Original plaintiffs find both success, disappointments. *The Advocate*, 1A.

Thornton, M. (1982, December 11). Administration offers alternative to court-ordered busing. *The Washington Post*, A9.

Tolbert, C. M. (1999). *Demographic findings on increasing residential desegregation and the resulting potential for public school desegregation in East Baton Rouge Parish, Louisiana*. Unpublished manuscript.

U.S. Bureau of the Census. (1999a). *Small Area Income and Poverty Estimates Program.* Washington, DC: U.S. Bureau of the Census.

U.S. Bureau of the Census. (1999b) *County population estimates for July 1, 1998 and population change for July 1, 1997 to July 1, 1998.* Population Estimates Program. Washington, DC: U.S. Bureau of the Census, Population Division.

U.S. Bureau of the Census. (1990). *Census of population and housing: Louisiana.* Washington: U.S. Government Printing Office.

U.S. Bureau of the Census. (1998). *Estimates of the population of counties by race and Hispanic origin:*

July 1, 1996. Population Estimates Program, Population Division, Washington, DC.

Wilson, W. J. 1987. *The truly disadvantaged.* Chicago: University of Chicago Press.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 2, 33-44

Science Achievement, Class Size, and Demographics: The Debate Continues

Marie Miller-Whitehead
Tennessee Valley Educators for Excellence

The purpose of the study was to examine the relationship between school system financial and demographic data and student achievement on the science section of the 1998 Tennessee state-wide Terra Nova tests to determine if systems which served large numbers of low income students had been as successful as more affluent systems in improving student achievement. The inquiry used categories from the system level data set from the Tennessee state report card, consisting of the Tennessee science scale scores for grades three through eight for 1998, per pupil expenditure, per capita income, percent of students on free and reduced lunch, percent ethnic, and percent of schools in system meeting state class size guidelines. Two separate analyses, one for mean scale scores by system across grades three through eight, and a second for grade five, were employed for this study. The first hierarchical regression analysis was conducted with variables regressed on student scale scores on the 1998 Tennessee state test administered to students in grades three through eight. The resultant equation yielded a Multiple R of .81 and an adjusted R^2 of .54, with percent of students on free and reduced lunch accounting for 54% of variability. Across grades three through eight, percent of students on free and reduced lunch accounted for the majority of explained variation in science scale scores. In the second procedure, for students in grade five, the same variables regressed on grade five science scale score achievement resulted in a formula which yielded a Multiple R of .81 and an adjusted R^2 of .53. At grade five, percent of schools meeting class size standards accounted for the greatest amount of explained variability in science scale scores. Per pupil expenditure had a positive effect on the science scale score achievement for grade five students.

Introduction[1]

In light of the enormous interest in student achievement, school performance indicators, and fiscal accountability that school systems face today, school system administrators must cope with an increasingly complex set of state and federal mandates and initiatives designed to insure that all student populations are provided the best education possible. School and school system accountability issues have also been influenced by funding equity lawsuits in many states, including the State of Tennessee, over the past ten years. The result has been restructuring of school system funding to help provide a more equitable access to excellence in education regardless of the county in which the students attend school.

This issue is made even more complex by accompanying initiatives aimed at increasing site-based decisions at the school level to assure that students receive an educational experience reflective not only of federal and state mandates, but of the communities in which they live. Thus, while some school systems may decide that a key objective is to improve student test scores on standardized tests, others systems may target improving student attendance, improving graduation rates, lowering dropout rates, or identifying other areas of achievement that they have identified as key to assuring that students are provided with every opportunity to become productive members of their communities, to continue their education, or to enter the workforce with the skills desired by 21st century employers. These skills may be very different from county to county, with rural, suburban, and urban employers often not having the same expectations for student graduates as those in neighboring counties. Additionally, systems may prioritize improvement, meaning that each student will show individual gain or growth in targeted skill areas; or, on the other hand, systems may set the goal that each student will achieve some predetermined mastery level on criterion referenced tests, thus setting a cut score all students are expected to attain. The State of Tennessee has mandated as one component of its legislated accountability system that all students will show improvement and gain in achievement on the state-wide assessments administered to all public school

---

Marie Miller-Whitehead is the director of the Tennessee Valley Educators for Excellence organization headquartered in Muscle Shoals, Alabama. Correspondence regarding this article should be addressed to Marie Miller-Whitehead, TVEE.ORG, P. O. Box 2882, Muscle Shoals, AL 35662 or by e-mail to marie@tvee.org.

[1] I gratefully acknowledge the invaluable assistance of Professor James E. McLean for his editorial suggestions, guidance, and for persevering with me through several revisions of this article. The current paper is based on an earlier version presented at the 1999 meeting of the Mid-South Educational Research Association in Point Clear, AL.

students in Tennessee each April in grades two through eight. Thus, while students in Tennessee take both criterion referenced and norm referenced tests in spring of each school year, at the state level systems are held accountable for student gains, or value-added, for each student who attends school in their system.

Given the variety of goals and objectives which school systems may choose to set for themselves, the present study of 1998 science scale score data for Tennessee school systems provides a focus on identifying those factors or variables that continue to effect student achievement on standardized tests. The results of the study are intended to provide school systems with information that will aid them in identifying schools and systems that can be benchmarked as setting a standard for achievement and excellence among their peers or school systems in comparable counties with comparable student populations.

Research Questions

This study is a continuation of a longitudinal meta-analysis of the Tennessee science scale score data beginning with 1991, now providing eight years of information for use by decision makers. Because the purpose of the study was to inform decision making and to identify variables that continue to effect student achievement subsequent to the restructuring of the state of Tennessee's school finance and accountability system under the Educational Improvement Act of 1991, the primary dependent variable for this study was scale scores, rather than the value added gain score computations that are designed to control for these variables. While school systems whose students are at the highest levels of achievement may focus on identifying other areas to improve while holding gains on test scores, most systems would certainly wish to include improvement in student academic achievement as a priority. Since under the value-added system each student is measured against his or her own performance on previous years' tests, the value-added computations are designed to control for student demographic variables such as individual ability, ethnicity, socioeconomic status, and school system characteristics that affect the rate at which students learn. All educators know that all students can learn, but not all will learn the same amount in the same length of time. Using the scale score data rather than the value-added data is an aid in accomplishing four tasks vital to school decision-makers:

1. How much are students learning compared to other students?
2. What are the factors that are affecting student learning in our system?
3. Are these factors, once identified, related to things which the school system can improve or change so that students achieve to the best of their ability?
4. Are other school systems more successful in helping their students achieve and if so, how?

While the strength of the value-added system lies in its ability to calculate student growth over time, regardless of confounding variables both within and outside of school system and student control (i.e., ability, health, family issues, gender, motivation and interest in school, socioeconomic factors, school climate, teaching, and leadership), scale scores are an additional and vital tool for school and community decision-makers, providing a quantitative basis for identifying those factors over which teachers, parents, school and system administration, and funding agencies exercise varying amounts of control.

Method

The initial reports in this longitudinal meta-analysis examined the aggregate school system data to determine differences in student science achievement by grade level and by year statewide for a five year period over grades two through eight (Miller-Whitehead, 1998a, 1999). Since the test instrument in use at the time, a version of the McGraw Hill CTBS4, had been normed in the late 1980s, statewide trends and patterns of achievement were more easily identifiable and not affected by changes in the test instrument, norm population, and the like, except for yearly production of the test drawn from items in the McGraw Hill test item pool. The results of these studies indicated that student achievement on the science subtest had shown gradual, but significant, improvement over a five year period, which confirmed the findings of earlier studies that had focused on providing evidence of validity and reliability of the TCAP (Tennessee Comprehensive Assessment Program) value-added model as one accountability indicator for Tennessee school systems. Subsequent studies focused on school level and system level data with the inclusion of student and system variables provided in the Tennessee School Report Card. Because it included data for all public schools and school systems in the state of Tennessee, this meta-analysis differs from other longitudinal studies which also make use of the Tennessee data, providing an extension of findings reported by other researchers. The Project STAR study has generated a wealth of information to assist policy-makers make decisions related to factors that affect student achievement in the primary grades and beyond but only includes classes and schools that are STAR Project participants and that are able to conform to STAR guidelines for class size and classes with and without teacher aides.

The school level mean scale scores for all grade levels at each school in each Tennessee system were used to compute system mean scale scores for grades 3 through 8 for a 5-year period using the MANOVA procedure with 5 levels for year and 6 levels for grade. While the original studies included data for grade 2, due to recent changes in state testing guidelines test score data is currently provided for grades 3 through 8. The data set consisted of 31 variables representing school level data for 138 Tennessee school systems. However, because several Tennessee school systems are not K-8 systems and thus did not have data for all grade levels included in the analysis, 133 systems entered the analysis for the omnibus MANOVA procedure. The decision was made not to replace missing values with the mean as the purpose of the study was to identify significant differences in student achievement from system to system. However, in the preliminary aggregation and examination of results by grade level all systems with grade level test score data entered the analysis for each tested grade level, with frequencies and descriptives for each school system.

This analysis yielded a mean scale score for each school system at each grade level for which student science test scores were available. These data were regressed on the following school system and county indicators made available on the Tennessee State Education Report Card: county per capita income, system per pupil expenditure, percent of students on free and reduced lunch, ethnicity, and percent of schools in system meeting state class size guidelines (Table 1). The 1996 data differed in several key areas from that of 1998; first that the test scale score data did not provide scores for second grade students, and second that beginning in spring of 1997 the CTBS4 was replaced with the CTBS5 Terra Nova, a new test more recently normed, with subsequently differing range of scale scores from those of the CTBS4. Published reviews of the most recent version indicate consistent reliability coefficients (.80 and .90 range) for the subtests (Monsaas, 1999; Nitko, 1999). The previous regression analysis conducted with the 1996 data had yielded an R^2 of .91 and an adjusted R^2 of .86 (Miller-Whitehead, 1998b, 1999). Of the variables in the analysis, percent free and reduced lunch was by far the most powerful predictor for school system science achievement across grade levels in 1996 ($r = -.94$, $p < .001$) with per capita income of the county in which the school system was located having a positive correlation ($r = .63$, $p < .05$) to system-wide student performance and per pupil expenditure ($r = .46$, $p < .1$) also having a positive correlation to achievement. The 1996 analysis did not, however, include the percent of schools in the system meeting Educational Improvement Act (EIA) class size guidelines which are defined by EIA and for the purposes of this study as a maximum class size for grades K-3 of 25, and an average of 20 with this number increased to a maximum of 30 for grades 4 through 6 and 35 for students in grades 7 and 8 (TCA 49-1-104). This variable was included in the present study because of initiatives and longitudinal studies which have shown class size to have a significant effect on student achievement (Achilles, 1999; Finn, 1998; Glass & Smith, 1978; Mosteller, 1995; Nye, Achilles, Boyd-Zaharias, Fulton, & Wallenhorst, 1994). Additionally, the financial burden placed on school systems due to state-mandated class size reduction initiatives pointed to the need for continuing examination of this variable as a correlate of student achievement (Grissmer, Flanagan, Kawata, & Williamson, 2000). As the purpose of the study was to benchmark characteristics of school systems which were among the highest levels of science achievement, results for the majority of the school systems with scores in the midrange were computed but not reported. Systems were rank ordered by science achievement and multiple analyses were conducted using systems randomly selected from among those in the upper and lower quartiles.

Review of the Literature

The Tennessee STAR Project studies and the Lasting Benefits Studies indicated that small classes do make a difference in student achievement, particularly in kindergarten through third grade, but that these differences may not last through the later grades unless the child has been in small classes for at least 3 or 4 consecutive years previously. The studies also indicated that students of lower socioeconomic status generally received more benefit from smaller class size than did students of higher socioeconomic status (Achilles, 1998; Achilles, Sharp, & Nye, 1998; Finn, 1998; Pate-Bain, Boyd-Zaharias, Cain, Word, & Binkley, 1997; Slavin, 1990). Studies of class size and student achievement yielded mixed results in South Carolina, Indiana, Texas, and Virginia, but these studies had various definitions for "small class size" and generally looked at reading achievement (Slavin, 1990). Systems that entered this analysis as having percent of schools meeting class size requirements may also have been STAR participants, but not necessarily so as the STAR guidelines for "small" classes require between 13-17 students per class in kindergarten through third grade (Achilles, 1999) where the state requires an average of 20 with no class over 25 (which would be classified as a "regular" class in the STAR Project). There is also considerable debate among researchers over the importance of differentiation between class size, the actual physical number of students in one teacher's classroom and pupil-teacher ratio (PTR) as the latter may include faculty and support personnel who are not actually in the classroom setting (librarians, guidance, principals), thus artificially distorting the numbers (Ferguson & Ladd, 1996; Pate-Bain et al., 1997; Achilles et al., 1998; Achilles, 1999). According to Achilles (1999), the difference between class size and pupil teacher ratio can be as much as 10 students per class and this is not a trivial matter. Thus, a reported PTR of 15:1 may actually mean that there are as many as 25 students in some if not all teachers' classrooms at any given school. Because the positive effects of small class size on student achievement through grades three and four had been well-documented in the literature earlier in the longitudinal STAR studies, this analysis examined students at grade 5 to determine if school systems, encouraged by positive results in previous years for kindergarten through third grade, were able to maintain the higher levels of achievement by meeting or lowering state class size guidelines for the cohort of students in fifth grade.

Of other variables in the analysis, ethnicity also deserves mention. Because of the various ways in which respondents self-report their ethnicity, it becomes more and more evident that for the sake of accuracy it is better to say or write that the results were obtained for "students who *identified themselves* as White, Black, Hispanic, Asian, or American Indian." Ethnicity and SES are aggregated to the school system level in this study. Schools nested within systems may have a more homogeneous ethnic and SES distribution due to county population patterns than at the system level. In this study, SES or socioeconomic status refers to percent of students in the school system entitled to the free and reduced lunch program according to guidelines established under ESEA Title I. While it is no secret that poverty has a negative effect on student achievement, Bryk and Raudenbush (1992) compared mathematics achievement of students in various public and parochial schools and determined that SES was less of a predictor of math achievement for students in Catholic schools than it was for students in public schools. Using NAEP science achievement data, Bruschi and Anderson (1994) found disparities in student science achievement by gender and ethnicity for students at ages 9, 13, and 17, with White students outperforming Hispanic and Black students (nearly a 55 point difference by age 17) and females performing better in nature science but males increasing with each grade level their early advantage in earth, space, and physical science. A study using the 1988 NELS data found that students began school with positive attitudes about learning math and science regardless of ethnicity or SES but that poor and minority students soon became discouraged and fell behind, and were thus not prepared to pursue careers in math and science (Huang, Peng, Salvucci, & Owings, 1995). We know from research on effective schools that some schools and school systems are more successful than others in helping all students achieve, thus SES was used in the present analysis in an effort to identify and benchmark school systems that had higher than predicted achievement.

Results

The range of scores for the 10 systems with the highest means was quite small (mean score A = 668.94 and mean score J = 661.35), while for the systems at the low end the range was greater (mean score P = 638.39 and mean score Z = 617.88). The range of scores for the entire sample (n = 21) was 51 scale score points (mean score A = 668.94, mean score Z = 617.88). Systems with the highest mean scores were more likely to be small systems of less than 5,000 students. However, systems that achieved the lowest mean scores represented both large and small systems. Of the systems that were among the highest in overall science scale scores, only three systems had more than the state average of students on free and reduced lunch (state = 41%, system B = 58%, system H = 45.5%, system I = 63.3%). For systems from among the lowest in science scale scores, all but one exceeded the state average for percent of students on free and reduced lunch.

The school systems in the sample represented K-12, K-9 or K-6 grade configurations. The system rank order of science scale score achievement yielded different results when examined for achievement at only specified grade levels (Table 2). School systems that appeared on both Table 1 (across all grades) and Table 2 (grade 5) are designated by the letter assigned on Table 1.

When aggregated at the grade level by system, several systems appeared consistently among the highest performing at several grade levels, including fifth grade science and across grades 3 through eight (systems A, D, and F). In systems with lower 8 achievement mean scores (systems Q, W, V, X, and Z), five of the systems were among the 10 lowest performing systems for grade 5 and also across grades 3 through 8. There was a greater range of scores for science achievement at fifth grade (mean scale score for system K= 684.25, mean scale score for system Z = 617.90, a range of 66 scale score points) than for science achievement across grades 3 through 8, with a range of 51 scale score points. While smaller systems once again achieved the highest science scores at grade 5, systems with lower mean science achievement scores included small, large, and very large school systems.

Table 1
System Descriptives High and Low Science Scale Scores Across Grades 3 to 8 (n = 21)

System	Science Mean Scale Score 1998	Per Pupil Expenditure 1997-98	% Schools Meeting Class Size	% Ethnic	% Free and Reduced 1997-98	County Per Capita Income 1997-98	System Size
A	668.94	$4,351	76.00%	92.7%	6.3%	27,888	15,000 to 25,000
B	667.35	$4,681	100.00%	99.5%	58.0%	12,287	1,000 to 5,000
C	667.32	$4,058	0%	82.4%	29.4%	14,485	less than 1,000
D	665.53	$3,769	100.00%	70.7%	34.2%	15,536	1,000 to 5,000
E	665.28	$4,331	100.00%	91.5%	30.5%	13,740	less than 1,000
F	664.03	$5,335	50.00%	93.7%	21.1%	17,042	1,000 to 5,000
G	664.00	$6,517	71.40%	83.6%	21.4%	19,056	1,000 to 5,000
H	663.32	$4,795	66.70%	92.5%	45.5%	18,161	1,000 to 5,000
I	661.97	$4,273	0%	96.1%	63.3%	15,024	less than 1,000
J	661.35	$4,073	100.00%	95.6%	33.1%	14,890	less than 1,000
P	638.39	$4,487	31.30%	66.0%	40.7%	21,216	35,000 to 50,000
Q	638.20	$3,847	62.50%	99.4%	60.2%	12,247	5,000 to 10,000
R	636.81	$3,774	90.00%	93.7%	48.4%	13,087	1,000 to 5,000
S	636.20	$4,862	52.20%	48.9%	49.8%	18,136	10,000 to 15,000
T	636.04	$4,425	87.50%	98.8%	59.6%	10,506	1,000 to 5,000
U	631.29	$3,777	85.70%	55.9%	52.7%	11,739	1,000 to 5,000
V	630.65	$4,826	100.00%	16.8%	87.5%	15,372	less than 1,000
W	630.48	$5,568	38.20%	50.4%	47.2%	23,804	50,000 to 75,000
X	623.16	$4,871	36.70%	13.3%	67.2%	21,486	75,000 +
Y	619.43	$4,836	100.00%	67.0%	65.2%	16,534	less than 1,000
Z	617.88	$3,988	37.50%	29.5%	82.8%	15,147	1,000 to 5,000
State TN	645.23	$4,391	56.30%	74.0%	41.0%	15,194	890,805

Table 2
Grade 5 Science Scale Score Variables (n = 20)

System	Science Mean Scale Score 1998 Grade five	Per Pupil Expenditure 1997-98	% Schools meeting Class Size	% Ethnic	% Free and Reduced 1997-98	County Per Capita Income 1997-98	System Size
K	684.25	$4,225	100.00%	98.6%	66%	13,908	1,000 to 5,000
L	675.60	$5,269	100.00%	98.0%	37%	19,056	1,000 to 5,000
D	673.90	$3,769	100.00%	70.7%	34%	15,536	1,000 to 5,000
A	673.65	$4,351	76.00%	92.7%	06%	27,888	15,000 to 25,000
F	670.38	$5,335	50.00%	93.7%	21%	17,042	1,000 to 5,000
G	665.60	$6,517	71.40%	83.6%	21%	19,056	1,000 to 5,000
M	665.40	$5,522	100.00%	92.7%	18%	27,888	1,000 to 5,000
N	664.25	$5,918	85.70%	98.5%	30%	14,591	1,000 to 5,000
O	664.00	$4,331	100.00%	91.5%	31%	13,740	less than 1,000
AA	663.93	$4,889	100.00%	94.9%	37%	11,300	1,000 to 5,000
AB	637.80	$3,665	50.00%	96.1%	34%	16,558	less than 1,000
AC	637.70	$3,651	0%	88.8%	50%	15,396	less than 1,000
AD	637.37	$3,767	100.00%	99.1%	38%	14,171	1,000 to 5,000
W	636.50	$5,568	38.20%	50.4%	47%	23,804	50,000 to 75,000
AE	634.30	$3,850	66.70%	85.9%	34%	12,500	1,000 to 5,000
Q	633.48	$3,847	62.50%	99.4%	60%	12,247	10,000 to 15,000
AF	632.00	$4,313	100.00%	44.2%	57%	16,558	1,000 to 5,000
V	631.00	$4,826	100.00%	16.8%	88%	15,372	1,000 to 5,000
X	630.78	$4,871	36.70%	13.3%	67%	21,486	75,000 +
Z	617.90	$3,988	37.50%	29.5%	83%	15,147	1,000 to 5,000
State	650.87	$4,391	56.30%	74.0%	41.0%	15,194	890,805

Table 3 provides school system level descriptive data, with range, minimum, maximum, mean scale score, standard error, standard deviation, and variance for each of grades 3 through 8 for the Tennessee science scale scores for the year 1998. Scores were recentered in 1998 with the introduction of the CTBS5 Terra Nova; the minimum mean score by system in grade 3 was 573, and the maximum mean score for grade 8 was 710. As expected, the range of scores for students in grade 3 was higher than for any other grade as students in their earliest school years have come from such diverse family and educational backgrounds and not all have acculturated equally to the expectations of the formal school setting.

Table 3
Descriptives by Grade Level for 1998 Tennessee Science Scale Scores, All Systems

Grade	N	Range	Minimum	Maximum	Mean	S.E.	Std	Variance
3	137	73.34	573	647	617	.96	11.2	126.07
4	137	61.66	599	660	634	.86	10.1	102.25
5	137	66.35	618	684	651	.80	9.3	86.62
6	136	55.83	625	681	660	.79	9.2	85.42
7	134	57.90	636	694	671	.83	9.6	92.34
8	134	65.95	644	710	685	.91	10.6	111.67

Science scale scores for 1998 and demographic data were examined to determine the relationship of the following variables on students' 1998 science achievement: percent of students on free and reduced lunch, percent of schools meeting state class size guidelines, county per capita income, system per pupil expenditure, and percent ethnicity of school system. The designation of percent of minority students has become somewhat problematic as in many school systems minority students comprise the majority of the student population. Hierarchical regression was determined to provide an approach which would account for both direct and indirect effects of the variables in the analysis and to allow a more parsimonious selection of variables. The selection of variables in hierarchical regression can be quite complex (Bryk & Raudenbush, 1992; Cook & Campbell, 1979; Mohr, 1995; Yates, 1996) and while many variables were available for inclusion in this study, the decision was made to limit them to those which the literature supported as having the greatest relationship to student achievement, in particular class size, poverty status, ethnicity, county per capita income, and per pupil expenditure. These data were downloaded electronically from the State of Tennessee website and matched with student achievement data either by county or by school system ID number.

A preliminary examination of correlations of the variables (Table 4) showed the strongest positive correlation (r = .69) across grades was associated with ethnic composition of the school system, while the strongest negative correlation (r = -.73) for percent of students qualified for free and reduced lunch. School systems in counties with a higher per capita income were slightly less likely to meet state class size guidelines than those in counties with lower per capita incomes (r = -.28), and systems in counties with higher per capita income were also more likely to have higher per pupil expenditures (r = .45). It also appeared likely that there were direct as well as indirect relations among the variables. Figure 1 provides a simplified graphic of the proposed path analysis model with the predicted direction of the direct effect of each variable on science scores across grades 3 through 8 and possible indirect effects of the variables.

Table 4
Pearson's Product Moment Correlations, Science Scale Scores Across Grades 3 through 8

	Scale Score	% at Class Size	% Ethnic	SES	Percapita$	Per-pupil$
scale score	--	.03	.69**	-.73**	.04	.06
% at class size	--	--	.15	-.02	-.28	-.09
% ethnic	--	--	--	-.56**	-.26	-.13
SES	--	--	--	--	-.39	-.18
percapita$	--	--	--	--	--	.45*
per-pupil$	--	--	--	--	--	--

Note. * $p < .05$, two-tailed; ** $p < .01$, two-tailed.

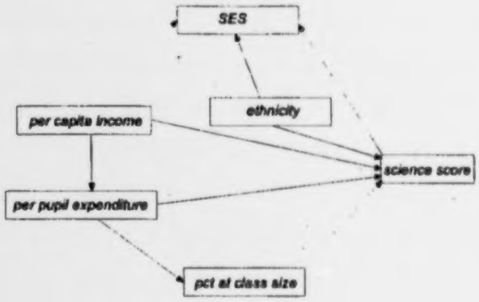


Figure 1. Initial path model showing expected influence of SES, ethnicity, per-pupil expenditure, percent meeting class size, and county per capita income on science achievement

Based on experience and student achievement research not only would poverty be expected to be negatively related to student test scores, there might also be an indirect relation of ethnicity and poverty to test scores if some ethnic groups are more likely than others to be poor or to qualify for entitlement programs based on family income (Anthony & Jacobson, 1992). It is often necessary to develop and test several models to be used in an analysis, and the preliminary attempts may represent a combination of research, experience, and wishful thinking. For this reason a graph such as is provided in Figure 1 may prove helpful in mapping combinations of relationships among the variables prior to conducting statistical tests. Figure 1 shows that SES is expected to have something of a negative effect on achievement, while county per capita income might be expected to have a positive direct effect on student achievement as well as a negative correlation to percent of students at poverty level. County per capita income was statistically significant when entered at step three of the regression analysis to test its indirect effect on per pupil expenditure and did have a positive effect for this variable across grades 3 through 8 (Table 5). A rather large disparity between county per capita income and percent of students at poverty level in the school system can be an indication of areas with high percentages of students enrolled in private schools or of areas where higher income residents do not have school age children.

A hierarchical regression entering all six variables in the first equation resulted in a Multiple R of .81 and an adjusted R^2 of .54. After adjusting for sample size (n = 21), these six variables accounted for 54% of variability in 1998 science scale scores across grades 3 through 8. Of the variables, only percent of students on free and reduced lunch was statistically significant (p = .033) at step one. However, as the number of cases per variable was less than that recommended for regression, the analysis may have underestimated the actual significance. There continues to be considerable debate over the issues of sample size and statistical significance (Chow, 1996; Cohen, 1987; Kaufman, 1998; Levin, 1998; McLean & Ernest, 1998; Thompson, 1998); however given the large effect size yielded by the variables in the regression, power for the cumulative R^2 at $\alpha = .05$ was determined to be .89 for the sample size of 21 (Stevens, 1990).

Table 5
Summary of Hierarchical Regression Analysis for Variables Predicting 1998 Science Scale Score Mean Across Grades 3 through 8

Variable	*B*	*SE B*	β
Step 1			
Percent free/reduced meals (SES)	-.52	.22	-.58*
Percent ethnic	.23	.15	.35
% Meeting class size	-.04	.09	-.07
Per pupil $	-.002	.005	.06
County per capita $	-.001	.001	-.14
Step 2			
Percent ethnic	-.41	.14	-.56**
Step 3			
Per-pupil $	.07	.032	.45*

Note. * $p < .05$, ** $p < .01$

The model resulted in the following equations, accounting for direct and indirect effects, for each of the dependent variables:

SS98:

$$\gamma' = 660.27 - .52(\%SES) + .23(\%ethnic) - .04(\%clssize) - .001(percapita\$) + .002(perpupil\$)$$

SES

$$\gamma' = 77.8 - .41(\%SES)$$

perpupil$:

$$\gamma' = 3363.55 + .07\ (percapita\$)$$

Systems A, B, H, and I had student science scale score achievement that exceeded predicted achievement, with system A and I having both higher student achievement and lower per pupil expenditure than predicted by the model. System B and H exceeded both predicted student achievement and per pupil expenditure, while system Z (a small system) had lower than predicted student achievement across grades 3 through 8 and lower system per pupil expenditure. The model was able to predict science scale score achievement accurately to within less than one standard deviation (18.32 points) and in several cases within 5 scale score points. Although not statistically significant, percent of schools in system within state class size guidelines had a slight negative effect on student science achievement over grades 3 through 8. Appendix B provides a graphical representation of the model that resulted from Figure 1. Because the literature strongly suggests that smaller class size has a positive effect on student achievement in the earlier grades (Achilles, 1999; Achilles & Price, 1999; Finn, 1998; Glass & Smith, 1978; Hanushek, 1998; Mosteller, 1995; Nye et al., 1994), a follow-up examination of the variables was conducted with grade 5 mean science scale score achievement as the dependent variable to determine if the class size variable had a different effect by grade level. As previously mentioned, results of Project STAR and other class size studies indicate that the greatest positive effect occurs in the earlier grades and a positive effect in later grades is largely dependent upon students having been in small classes (13 to 17 students) for 3 to 4 consecutive years. The results of preliminary examination of correlation of the variables is provided in Table 6. Percent of students on free and reduced lunch showed a negative correlation to grade 5 science achievement (r = -.61) while percent of classes meeting state class size guidelines showed a positive correlation to grade five science achievement (r = .49). Ethnicity and SES were negatively correlated and appeared to be multicollinear (r

= -.66) but as both variables were of interest both were used in the analysis.

Table 6
Pearson's Product Moment Correlations for Grade 5 Science

	scale score 98.5	% at class size	% ethnic	SES	percapita$	per-pupil$
scale score 98.5	--	.49**	.57**	-.61**	.19	.36
% at class size	--	--	.19	-.17	-.10	.13*
% ethnic	--	--	--	-.66**	-.20	-.09
SES	--	--	--	--	-.61	-.29
percapita$	--	--	--	--	--	.41*
per-pupil$	--	--	--	--	--	--

Note. * $p < .05$, two-tailed, ** $p < .01$, two-tailed

A hierarchical regression entering the six variables in the first equation yielded a Multiple *R* of .81 and an adjusted R^2 of .53. After adjusting for sample size (n = 20), the five variables accounted for 53% of variability in the 1998 science scale score achievement of grade 5 students. Of the variables in the analysis, only class size was statistically significant (p = .043), accounting for 37% of the explained variability in grade five science achievement. Science scale scores increased .25 point for each percent increase in schools which met state class size guidelines. A stepwise regression procedure with the same variables resulted in the removal of all but two variables (percent of students on free and reduced lunch and class size) which yielded an adjusted R^2 of .47, less than the variability accounted for by the entry method. For a sample size of 20 at an alpha of .05, power for the analysis was determined to be .82.

Table 7
Summary of Hierarchical Regression for Variables Predicting Grade 5 Science Scale Score Achievement

Variable	*B*	*SE B*	β
Step 1			
Percent free and reduced lunch (SES)	-.802	28.10	-.009
Percent ethnic	40.10	19.79	.568
% Meeting class size	.25	.11	.368*
Per pupil $	.006	.004	.269
County per capita $	.0009	.001	.224
Step 2			
Percent ethnic	-.58	.103	-.76***
County per capita $	-.000024	.000	-.52**

Note. * $p > .05$, ** $p > .01$, *** $p > .001$

To determine the indirect effect of the variables on science achievement, percent of students on free and reduced lunch was regressed on percent white students (non-minority) and county per capita income, and the resultant equation yielded a Multiple *R* of .84 and an adjusted R^2 of .66. While the indirect effect of variables on class size was of interest, the variables in the present analysis provided insignificant R^2 changes when regressed on class size and thus were not included in the model.

The model yielded the following equations for the variables which entered each step of the analysis:

SS98.5:

$$\gamma' = 557.39 - .80(\text{SES}) + .25(\text{class size}) + 40(\text{ethnic}) + \\ .006(\text{perpupil\$}) + .0009(\text{percapita\$})$$

SES

$$\gamma' = 1.27 - .58(\text{ethnic}) - .00002(\text{percapita\$})$$

The systems for which variables for percent ethnic and percent free and reduced lunch were nearer the mean (i.e., system D) exhibited much closer model fit before controlling for indirect effects than those systems which varied most from the mean (System K, A, W, Z) for ethnicity and poverty. Before controlling for indirect effects, systems W, X, and Z had student achievement results considerably higher than predicted by the model, based on values for the variables used in the equation. The underestimation of the model may be indicative of a curvilinear relation for some of the variables and transformation of these may have resulted in better model fit; however, for ease of interpretation no variables were transformed for the present study.

Discussion

Tennessee provides each school system with value-added scores based on computation of student performance over a 3-year period, the results of which provide a value-added gain score of teacher, school, and school system performance. This procedure thus controls for the confounding variables of student ability, ethnicity, socioeconomic status, school efficacy, community involvement and support, class size, and the myriad other confounding variables which affect student achievement. The present study provides a one year single point in time analysis with the purpose of determining at which grade levels and for which school systems variables continue to play a significant role in student achievement on the science subtest of the statewide test. Therefore, results may differ from subtest to subtest, and reading, language arts, social studies or math scale scores might not have resulted in the same equation for the variables. While the rationale for using scale scores as a measure is to provide a method for comparison of scores on the same subtest, this does not apply to comparisons across subtests. The variable for class size may well have a different effect for reading and language arts depending upon grade level than the effect on the science results.

Thus, while across grade levels 3 through 8 the results of this study indicate that smaller class size was not significant in increasing student science achievement, at grade 5, school systems which were at state recommended class size realized an increase of up to 25 scale score points on the science subtest. This is an indication that class size continued to have a positive effect on student test achievement through grade 5. It is also important to note that science scale score achievement of students in grades 6, 7, and 8 (the across grade data) only accounted for percent of school systems meeting class size requirements in 1998 and thus did not take into account the possible effect on student achievement of class size or other variables in prior years (Sanders & Rivers, 1996). In addition, while maximum class size for grades K-3 is 25, and an average of 20, this number increased to a maximum of 30 for grades 4 through 6 and 35 for students in grades 7 and 8 (TCA 49-1-104). This study, as a meta-analysis, also did not account for the possibility that many of the higher performing systems reported to have been 100% within state class size guidelines may in actuality have had classes considerably smaller than required. Thus, for grade 5, percent of classes meeting class size requirements may not tell the entire story as the grade 5 classes in the analysis may have had any number of students less than 30. Though class size recommendations are based on many more factors than can be measured by a standardized test (French, 1993), meeting or lowering state class size guidelines is certainly indicated for those schools concerned with improving student achievement, as this one variable is well within the control of a school system and can help to offset disadvantages due to student poverty and ethnicity so that all students have an opportunity to succeed in school. Several recently released reports by the STAR project leaders and RAND provide a cost analysis overview related to implementation of student achievement initiatives including smaller class size (Achilles, Finn, Gerber, Pannozzo, & Boyd-Zaharias, 2000; Grantham, 2000; Achilles & Price, 1999; Grissmer et al., 2000). Because there are many questions which cannot be answered by looking at the numbers, the results of this study would seem to call for follow-up by telephone, paper survey, or site visits to the high-performing systems to provide a more personal glimpse of "how they do it." Many systems have implemented comprehensive inservice programs to promote action research within the classroom, thus providing teachers, administrators, and support personnel with consistent tools for measurable results (McLean, 1995).

For grade 5 students, class size accounted for 37% of the explained variability in science scale scores. This finding can be considered quite significant based on the results of other studies which have failed to find that class size makes a difference past the first two or three years of school. However, follow-up studies on Project STAR student participants have shown positive effects continuing at least into grade 9 (Achilles, 1998; Pate-Bain et al., 1997). Because the present study is a meta-analysis of statewide data, no attempt was made to determine whether the highest performing systems were also participants of projects such as the STAR project where researchers examine results for individual classrooms of students in strictly controlled experimental conditions. The percent of students on free and reduced lunch had little direct effect on scale scores at grade 5. This finding is indicative that Tennessee school systems had implemented effective programs for improving student science achievement of low income students in grade five. Across grades 3 through 8, percent of students on free and reduced lunch accounted for 58% of explained variation in student science achievement, with ethnic composition of the school system having a direct effect of 35% as well as an indirect effect by meal status. Although across grades 3 through 8, per pupil expenditure had little direct effect on student science achievement (6%), per pupil expenditure accounted for 27% of explained variability at grade 5. Although there are many factors beyond the scope of this study which comprise the large unexplained variability in student science achievement such as school leadership, teaching, school climate, home environment and the like, money invested in education did make a difference. County per capita income did not appear to have either a direct or indirect effect on student science achievement at the fifth grade level, an indication that funding equity initiatives across the state were successful through grade 5 for science achievement.

From the perspective of the effective schools movement (Lezotte, 1989; 1993) the findings indicated that while many Tennessee systems had science scale score achievement higher than expected based on system demographics, others (those with extreme values on the variables) should examine a variety of alternatives to improve opportunities to learn so that their student achievement is more evenly distributed across student groups, i.e., by ethnicity, gender, and socioeconomic status.

References

Achilles, C. M. (1998). *Small-class research supports what we all know (So, why aren't we doing it?).* (ERIC Document Reproduction Service No. ED 419 289)

Achilles, C. M. (1999). *Let's put kids first, finally: Getting class size right.* Thousand Oaks, CA: Corwin Press.

Achilles, C. M., & Finn, J. D. (1999). *Some connections between class size and student achievement.* (ERIC Document Reproduction Service No. ED 430 943)

Achilles, C. M., Finn, J. D., Gerber, S. B., Pannozzo, G., & Boyd-Zaharias, J. (2000, October). *Small classes do reduce the test score achievement gap.* Paper presented at the Council of Great City Schools Conference, Los Angeles, CA.

Achilles, C. M., & Price, W. J. (1999). Can your district afford smaller classes in grades K-3? Can smaller classes be cost effective? *School Business Affairs, 65*(1), 10-16.

Achilles, C. M., Sharp, Mark, & Nye, B. A. (1998). Attempting to understand the class size and pupil-teacher ratio (PTR) confusion: A pilot study. (ERIC Document Reproduction Service No. ED 419 290)

Anthony, P. & Jacobson, S. (1992). *Helping at-risk students: What are the educational and financial costs?* Newbury Park, CA: Corwin Press.

Bruschi, B. A., & Anderson, B. T. (1994). *Gender and ethnic differences in science achievement of nine-, thirteen-, and seventeen-year-old students.* (ERIC Document Service No. ED 382 751)

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage Publications.

Chow, S. L. (1996). *Statistical significance: Rationale, validity, and utility.* London: Sage Publications.

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin Company.

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education.* Washington, DC: The Brookings Institution.

Finn, J. D. (1998). *Class size and students at risk: What Is known? What is next?* Washington, D. C: U.S. Department of Education. Office of Educational Research and Improvement. (AR 98 7104)

French, N. K. (1993). Elementary teacher stress and class size. *Journal of Research and Development in Education, 26*(2), 66-73.

Glass, G. V., & Smith, M. L. (1978). *Meta-analysis of research on the relationship of class size and achievement.* San Francisco: Far West Laboratory for Educational Research and Development.

Grantham, M. K. (2000, November). Impact of small class size on achievement. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Bowling Green, KY.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us.* Santa Monica: RAND.

Hanushek, E. A. (1998). *The evidence on class size.* Rochester, NY: The University of Rochester, W. Allen Wallis Institute.

Huang, G., Peng, S., Salvucci, S., & Owings, J. (1995). *Understanding racial-ethnic differences in secondary school science and mathematics achievement. Research and development report.* (ERIC Document Reproduction Service No. ED 381 342)

Kaufman, A. S. (1998). Introduction to the special issue on statistical significance testing. *Research in the Schools, 5*(2), 1-3.

Levin, J. R. (1998). What if there were no more bickering about statistical significance tests? *Research in the Schools, 5*(2), 43-54.

Lezotte, L. W. (1989). Base school improvement on what we know about effective schools. *American School Board Journal, 176*(8), 18-20.

Lezotte, L. W. (1993). Creating effective schools today and tomorrow. *Journal for Quality and Participation, 16*(1), 22-30.

McLean, J. E., & Ernest, J. M. (1998). The role of statistical significance testing in educational research. *Research in the Schools, 5*(2), 15-22.

McLean, J. E. (1995). *Improving education through action research: A guide for administrators and teachers.* Thousand Oaks, CA: Corwin Press, Inc.

Miller-Whitehead, M. (1998a). *An analysis of science scale scores for grades 2-8 in Tennessee for 1990 to1994* (ERIC Document Reproduction Service No. ED 414 160)

Miller-Whitehead, M. (1998b). *A longitudinal analysis of science scale scores grades 2-8 in Tennessee for 1992-1996* (ERIC Document Reproduction Service No. ED 414 170)

Miller-Whitehead, M. (1999). *Tennessee TCAP science scale scores 1990-1997: Implications for continuous improvement and educational reform* (ERIC Document Reproduction Service No. ED 416 070)

Mohr, L. B. (1995). *Impact analysis for program evaluation.* Thousand Oaks, CA: Sage Publications.

Monsaas, J. A. (1999). Review of the Terra Nova (Comprehensive Test of Basic Skills, Fifth Edition). In B. S. Plake & J. C. Impara, (Eds.). *The supplement to the thirteenth mental measurements yearbook.* Lincoln: University of Nebraska, Buros Institute of Mental Measurements.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2), 113-127.

Nitko, A. J. (1999). Review of the Terra Nova (Comprehensive Test of Basic Skills, Fifth Edition). In B. S. Plake & J. C. Impara, (Eds.). *The supplement to the thirteenth mental measurements yearbook.* Lincoln: University of Nebraska, Buros Institute of Mental Measurements.

Nye, B. A., Achilles, C. M., Boyd-Zaharias, J., Fulton, B. D., & Wallenhorst, M. (1994). Small is far better. *Research in the Schools, 1*(1), 9-20.

Pate-Bain, H., Boyd-Zaharias, J., Cain, V., Word, E., & Binkley, M. (1997). *STAR Follow-up studies, 1996-1997: The student/teacher ratio (STAR) Project.* (ERIC Document Reproduction Service No. ED 419 593)

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville: The University of Tennessee Value-Added Research and Assessment Center.

Slavin, R. E. (1990). School and classroom organization in beginning reading: Class size, aides, and instructional grouping. In R. E. Slavin, N. L. Karweit, and B. A. Wasik (Eds.), *Preventing early school failure: Research, policy, and practice.* (pp. 122-142). Boston: Allyn and Bacon.

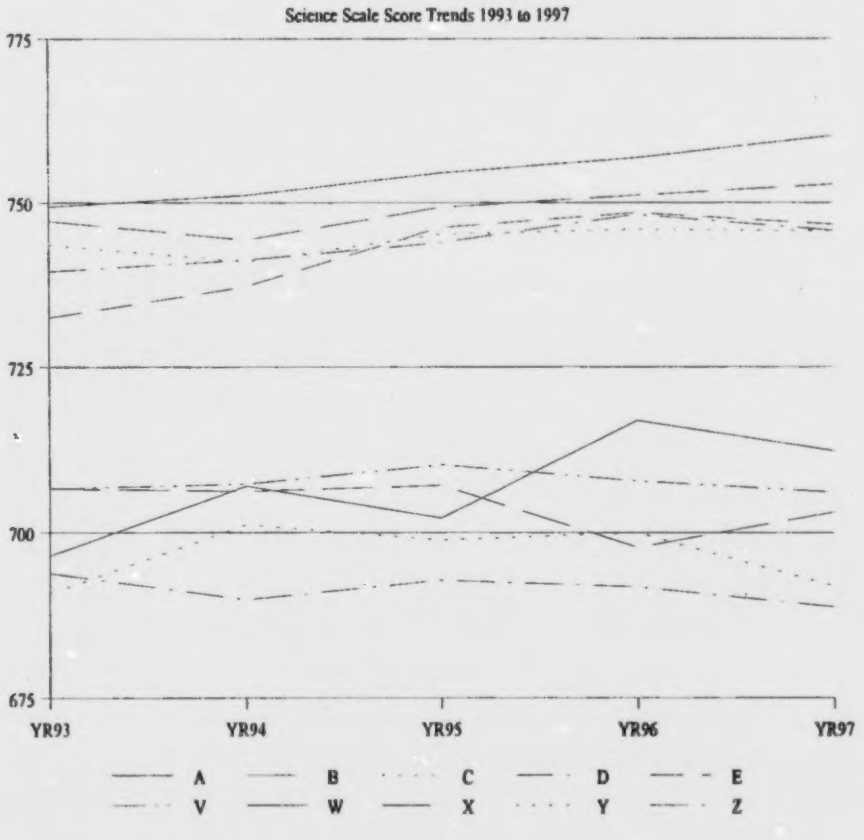
Stevens, J. (1990). *Intermediate statistics: A modern approach.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. *Research in the Schools, 5*(2), 33-38.

Yates, B. T. (1996). *Analyzing costs, procedures, processes, and outcomes in human services.* Thousand Oaks, CA: Sage Publications.

APPENDIX A

Five Years Trends of System Means for High and Low Achieving Tennessee Systems



APPENDIX B
Direct and Indirect Effects of Variables on Science Achievement Across Grades 3 - 8

$B_{\text{science SES}} = -.58$

$B_{\text{science ethnic}} = .35$

$B_{\text{science cls size}} = -.07$

$B_{\text{science perpupil}} = .06$

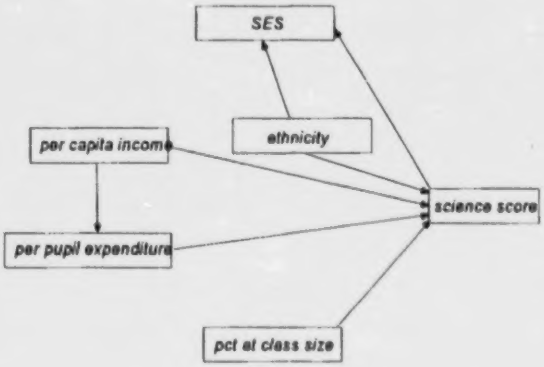
$B_{\text{science percapita}} = -.14$

$B_{\text{ethnicSES}} = -.56$

$B_{\text{percapita perpupil}} = .45$

$B_{\text{ethnicSES *science}} = -.20$

$B_{\text{percap perpupil *science}} = .03$



Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 2, 45-57

Characteristics of Effective Teachers: Perceptions of Preservice Teachers

Ann E. Witcher
University of Central Arkansas

Anthony J. Onwuegbuzie
Howard University

Lynn C. Minor
Valdosta State University

The purpose of this study was to determine preservice teachers' perceptions about the characteristics of effective teachers, as well as to investigate factors (e.g., gender, ethnicity, age, year of study, area of specialization, and parental status) that may have influenced their responses. Participants were 219 students attending a large mid-southern university. These students were administered a questionnaire asking them to identify, to rank, and to define between 3 and 6 characteristics that they believed excellent teachers possess or demonstrate. A phenomenological analysis (i.e., method of constant comparison) of responses revealed several characteristics that many of the preservice teachers considered to reflect effective teaching. In order of endorsement level, the following six themes emerged from these characteristics: (a) student-centeredness (79.5%), (b) enthusiasm for teaching (40.2%), (c) ethicalness (38.8%), (d) classroom and behavior management (33.3%), (e) teaching methodology (32.4%), and (f) knowledge of subject (31.5%). A canonical correlation analysis revealed that females, college-level juniors, and minority students tended to endorse teacher characteristics that were associated with ethical behavior and teaching methodology to a greater extent than did their counterparts. They also tended to rate attributes that were associated with knowledge of subject and classroom and behavior management to a lesser degree. Age served as a suppressor variable. Using ipsative/cluster analyses, four profiles of students' responses to the six themes emerged. The implications of these findings are discussed, as are recommendations for future research.

Throughout the 20th Century, there have been continuing attempts to identify characteristics of effective teachers. Currently, most definitions reflect the notion that effectiveness is determined by using the parameters of classroom instruction. Thus, effective teachers are generally described as business-like in teaching, clear and specific in the use of language, and adept in the use of paralanguage. They sequence and schedule lessons that include detailed explanations and examples, provide immediate and corrective feedback, and ensure plenty of practice time. One need not look far into the literature to find further characteristics of effective teachers as identified through research. What follows is a sampling of that literature.

Good and Brophy (1994) described effective teachers as active teachers who make maximum use of instructional time, present material in ways to meet student needs, monitor programs and progress, and plan opportunities for students to apply newly acquired concepts and skills. These teachers also re-teach when needed, maintain high but realistic goals, and provide motivation when introducing material both during and at the conclusion of lessons.

Berliner (1985) listed the following 10 characteristic behaviors of effective teachers: (a) monitor students' work, check progress, and give feedback; (b) structure lessons and communicate expectations through objectives; (c) pace instruction rapidly; (d) ask questions requiring analysis, synthesis, and evaluation; (e) provide wait time for answers; (f) communicate high expectations; (g) provide a safe and orderly classroom; (h) foster a convivial atmosphere; (I) make the best motivational use of tests and grades; and (j) provide feedback in positive, corrective ways.

Effective teachers also have been described as those who encourage active student participation and make relevant assignments, arrange for plenty of successful engaged time, are skillful in using questions, and employ

---

Ann E. Witcher is an associate professor in the Department of Curriculum and Instruction at the University of Central Arkansas. Anthony J. Onwuegbuzie is an associate professor in the Department of Human Development and Psychoeducational Studies at Howard University. Lynn Minor is a doctoral candidate in Educational Leadership at Valdosta State University. Correspondence should be addressed to: Ann E. Witcher, Department of Curriculum and Instruction, 217 Mashburn Hall, University of Central Arkansas, Conway, AR 72035-0001 or by e-mail to annw@mail.uca.edu.

the use of wait-time when seeking student response (Finn, 1993; Good & Brophy, 1994; Redfield & Rousseau, 1981; Rosenshine & Stevens, 1986; Tobin, 1987).

In describing instructors who are effective, Wortruba and Wright (1975) provided the following characteristics: knowledgeable about and enthusiastic in presenting subject matter, organized in lesson presentation, flexible in using a variety of teaching strategies, effective in communication, positive in attitude toward students, and fair in assessment and grading procedures.

According to Cotton (1995), teachers are effective when they pre-plan curricula and integrate traditional school subjects where appropriate, provide clear expectations for students, carefully orient students to lessons, and are clear and focused in instruction. Apparently, effective teachers provide feedback and reinforcement, review and re-teach when needed, use questions effectively, monitor student progress, and use both traditional and alternative assessment procedures. Additionally, group designs meet student academic and affective needs, and there is efficiency in the use of instructional time and in the running of the classroom. Both critical and creative thinking are promoted, and workplace readiness skills are integrated into subject matter. Finally, these teachers provide incentives through recognition and rewards, display positive interactions with their students, and are consistent and equitable in their treatment of students.

A description of effective teachers as being strong in student-teacher relationships is offered by Wubbels, Levy, and Brekelmans (1997). Believing that solid student-teacher relationships are the very foundation for a positive classroom climate, these researchers posit that effective teachers are those who are flexible in their abilities to be both dominant and cooperative, empathetic yet in control. Teachers who are effective allow for pupil freedom and responsibility, and they reflect on student feedback so that their views of self closely resemble the perceptions of students. They are skilled in analyzing student needs for relationship behavior, and they are adept in meeting those needs.

Reed and Bergemann (1992) identified effective teachers as possessing skills that may be characterized as "measurable" and "not measurable." Examples of the 37 measurable skills include knowledge of subject matter, effective and efficient use of time, clear communication, and organization. There are 13 non-measurable skills, and these are exemplified by qualities such as sound moral character, sensitivity, warmth, and patience.

Norton (1997), who conducted semi-structured interviews with first-year elementary teachers (n = 42), identified characteristics of effective teachers. For this group, the effective teacher was profiled as caring and committed, demonstrating genuine affection and respect for students and about students and teaching; highly creative, which included originality, flexibility, and fluency; providing a stimulating classroom environment; proficient in reflective thinking; responsive to unique educational and emotional needs of students; constantly reviewing instructional goals, methods, and materials; and possessing a strong internal locus of control--that is, willing to do anything needed to help a student improve.

One view of effective teaching characteristics is provided by *Pathwise: A Framework for Teaching*, developed by Educational Testing Services (1997) as part of its Praxis Series. This framework identifies four domains: planning and preparation, classroom environment, instruction, and professional responsibilities. Within these domains are 19 skills necessary for effective teaching. These include demonstrating knowledge of content and pedagogy, creating a climate of respect and rapport, clearly and accurately communicating, and reflecting on teaching.

According to Roueche, Baker, Mullin, and Boy (1986), characteristics of effective teachers can be grouped into three categories: motivation, interpersonal skills, and cognitive skills. Motivation encompasses a strong commitment to students and learning, goal orientation, integrated perception, and reward orientation. Interpersonal skills are reflected in objectivity, active listening, rapport, and empathy. The category of cognitive skills includes individualized perception, teaching strategies, knowledge, and innovation.

The American Association of School Administrators (AASA) researched characteristics of effective teachers and reported that qualities tend to fall into two categories: (a) management and instructional techniques and (b) personal characteristics (Demmon-Berger, 1986). Generally speaking, the AASA describes effective teachers as good managers who handle discipline through prevention; use systematic, yet varied, instructional techniques; are knowledgeable of subject matter and task oriented, while tailoring teaching to student needs; are highly flexible, enthusiastic, and imaginative and emphasize perceptual meanings more than facts and events; believe in their own abilities and have high expectations; are democratic in their approach and display warmth, care, and concern when interacting with students; and are readily accessible outside of class.

The National Board for Professional Teaching Standards (NBPTS) was established in 1987 to strengthen the teaching profession and thereby improve learning. This board attempts to identify and to recognize teachers who effectively promote student learning and who demonstrate high levels of knowledge, skills, dispositions, and commitments, as reflected in the following five core propositions: (a) teachers are committed to students and their learning; (b) teachers have extensive knowledge about the subjects they teach and how to teach these subjects to their students; (c) teachers are responsible for managing and monitoring student learning; (d) teachers reflect on their practice and learn from their experiences; and (e) teachers are members of learning communities. According to the NBPTS, these five elements form the core attributes of an effective teacher (NBPTS, 1987).

Although the literature abounds with information regarding teacher effectiveness, the majority of these articles do not represent primary studies. Of the formal investigations undertaken in this area, most have examined actual characteristics of effective teachers or have asked inservice teachers and educational theorists about their beliefs regarding effective teaching; that is, relatively few researchers have studied the *perceptions* of preservice teachers concerning the attributions of effective teachers. Moreover, most of the investigations have utilized qualitative techniques (e.g., interview) using small samples. A paucity of studies have incorporated qualitative and quantitative analyses within the same framework. This was the goal of the present investigation. Specifically, the purpose of this study was to investigate what pre-service teachers view as important characteristics of effective teachers, with the intent of comparing their responses to descriptions provided in the literature. Also of interest was to investigate factors (e.g., gender, ethnicity, age, year of study, area of specialization, and parental status) that may have influenced their responses. It was hoped that findings from this study would help educators to determine the extent to which the perceptions of preservice teachers are similar to those of more experienced individuals.

Method

Participants

Participants were 219 preservice teachers who were attending a mid-southern university. Using Miles and Huberman's (1994) typology of sampling strategies in qualitative inquiries, the type of sampling incorporated in the present investigation was *criterion*, with the eligibility criterion being that the sample members were preservice teachers with no teaching experience. The majority of the sample was female (72.1%). With respect to ethnicity, the respondents comprised Caucasian-American (87.1%), African-American (10.0%), Hispanic (1.0%), Asian-American (0.5%), and other (1.5%). Ages ranged from 19 to 50 ($M = 24.2$, $SD = 6.1$). With regard to year of study, participants were either juniors (46.0%), seniors (45.5%), or post-baccalaureate (8.4%). Nearly all students (94.7%) had attended a public high school, with the location of their schools being predominantly in either a suburban (39.6%) or a rural (43.7%) setting. Consistent with their backgrounds, the majority of students intended to teach either at a public-suburban school (39.5%) or at a public rural school (32.3%).

Instruments and Procedures

Participants were administered a questionnaire during class sessions asking them to identify, to rank, and to define between 3 and 6 characteristics that they believed excellent teachers possess or demonstrate. This questionnaire also extracted the following demographic information: gender, age, major, year of study (i.e., junior vs. senior vs. post-baccalaureate), ethnicity, type of high school attended by respondent (i.e., public vs. private), location of high school attended (i.e., suburban vs. urban vs. rural), type of school in which the student would most like to teach (i.e., public-urban vs. public-suburban vs. public-rural vs. private-church sponsored vs. private-non church sponsored), and whether the respondent was a parent of a school-aged child. *Reciprocity* (Creswell, 1998) was established between the researchers and sample members inasmuch as participants received extra course credit.

Analysis

A sequential mixed-methodological analysis (SMMA), as described by Onwuegbuzie (2000), was undertaken to analyze the data. This analysis involved the qualitative and quantitative data analytic techniques in a sequential manner, commencing with qualitative analyses, followed by quantitative analyses that built on the qualitative analyses, and then ending with qualitative analyses. Utilizing the framework of Greene, Caracelli, and Graham (1989), the purpose of the mixed-methodological analysis was development, that is, using the methods sequentially such that results from one data-analytic method inform the use of the other method. More specifically, the goal of the SMMA was typology development (Caracelli & Greene, 1993).

The SMMA involved five stages. The first stage consisted of a phenomenological mode of inquiry (i.e., exploratory stage) to examine the responses of students regarding their perceptions of characteristics of effective teachers (Goetz & Lecompte, 1984). The phenomenological method essentially represents an attempt to understand phenomena from the perspective of those being studied (Goetz & Lecompte, 1984). Phenomenological analyses are inductive, generative, and constructive because they require the researcher(s) to *bracket* or suspend all preconceptions (i.e., *epoche*) in order to avoid unduly biasing the analyses (Moustakas, 1994). Thus, the researchers were careful not to form any *a priori* hypotheses with respect to preservice teachers' perceptions of effective teacher characteristics.

The phenomenological analysis undertaken in the present investigation involved the methodology of *reduction* (Creswell, 1998). Specifically, a modification of Colaizzi's (1978) phenomenological analytic methodology was utilized. The procedural steps used were as follows: (a) all the participants' descriptions, phrases, and sentences were read in order to acquire a feeling for them; (b) these participants' responses were then *unitized* (Glaser & Strauss, 1967); (c) these units of information served as the basis for extracting a list of nonrepetitive, nonoverlapping significant statements (i.e., *horizonalization* of data), with each statement treated as having equal worth. Units were eliminated that contained the same or nearly the same statements such that each unit corresponded to a unique teacher characteristic; (d) meanings were formulated by specifying the meaning of each significant statement (i.e., unit); and (e) clusters of themes were organized from the aggregate formulated meanings, with each cluster containing units that appeared similar in content such that each cluster represented a distinct emergent theme (i.e., *method of constant comparison*; Glaser & Strauss, 1967; Lincoln & Guba, 1985). These clusters of themes were referred back to the original descriptions in order to validate them. This was undertaken in order to ensure that no original descriptions were unaccounted for by the cluster of themes and that no cluster contained units that were not in the original descriptions. The themes were created *a posteriori* (Constas, 1992).

This five-step method of analysis was utilized to reveal a number of themes relating to students' perceptions of characteristics of effective teachers. The locus of typology development was *investigative*, emerging from the intellectual constructions of the researchers; similarly, the source for naming of categories was investigative (Constas, 1992). *Double coding* (Miles & Huberman, 1994) was used for categorization verification in the form of inter-rater reliability. As such, the verification component of categorization was *empirical* (Constas, 1992). Specifically, two of the researchers independently coded the respondents' descriptions and determined the emergent themes. The themes and classification themes were compared and rate of agreement assessed (i.e., inter-rater reliability). Because a quantitative technique (i.e., inter-rater reliability) was utilized as a validation technique, in addition to being empirical, the verification component of categorization was *technical* and the verification approach was accomplished *a posteriori* (Constas, 1992).

The second stage of the mixed-methodological analysis involved utilizing descriptive statistics (i.e., exploratory stage) to analyze the hierarchical structure of the emergent themes (Onwuegbuzie, 2001; Onwuegbuzie & Teddlie, in press). In particular, each theme was *quantitized* (Tashakkori & Teddlie, 1998). Specifically, for each participant, a score of "1" was given for a theme if it represented at least one of the six stated characteristics; otherwise, a score of "0" was given for that theme. That is, for each sample member, each theme was quantitized either to a score of "1" or a "0" depending on whether it was represented by that individual. This dichomotization led to the formation of an *inter-respondent matrix* (i.e., *participant x theme matrix*) and a *intra-respondent matrix* (i.e., *unit x theme matrix*) (Onwuegbuzie, 2001; Onwuegbuzie & Teddlie, in press). Both matrices contained a combination of 0s and 1s. The quantitizing of themes allowed the computation of what Onwuegbuzie (2001) termed as *manifest effect sizes* (i.e., effect sizes pertaining to observable content). The following two types of manifest effect sizes were computed. First, by calculating the frequency of each theme from the inter-respondent matrix, percentages were computed to determine the prevalence rate of each theme. The latter served as *frequency effect size* measure (Onwuegbuzie, 2001; Onwuegbuzie & Teddlie, in press). Second, the intensity of each identified theme was determined via the intra-respondent matrix. Specifically, the number of characteristics cited for each theme and, consequently, the proportion of characteristics identified per theme. The latter served as an *intensity effect size* measure (Onwuegbuzie, 2001; Onwuegbuzie & Teddlie, in press).

The third stage of the mixed-methodological analysis involved the utilization of the inter-respondent matrix to conduct an exploratory factor analysis to ascertain the underlying structure of these themes (i.e., exploratory stage). This factor analysis determined the number of factors underlying the themes. These factors, or latent constructs, represented *meta-themes* (Onwuegbuzie, 2001; Onwuegbuzie & Teddlie, in press) such that each meta-theme contained one or more of the emergent themes. The *trace*, or proportion of variance explained by each factor after rotation, served as a *latent* effect size for each meta-theme (Onwuegbuzie, 2001; Onwuegbuzie & Teddlie, in press). Additionally, a manifest effect size was computed for each meta-theme by determining the combined frequency effect size for themes within each meta-theme (Onwuegbuzie, 2001; Onwuegbuzie & Teddlie, in press). By determining the hierarchical relationship among the themes, in addition to being empirical and technical, the verification component of categorization was *rational* (Constas, 1992).

The fourth stage of the mixed-methodological analysis involved the determination of antecedent correlates of the emergent themes that were extracted in Stage 1 and quantitized in Stage 2 (i.e., confirmatory analyses). This phase utilized the inter-respondent matrix to undertake (a) a series of Fisher's Exact tests to determine which background variables were related to each of the themes; and (b) a canonical correlation analysis to examine simultaneously the relationship between the themes and the demographic variables. A canonical correlation analysis is used to determine the relationship between two sets of variables when each set contains more than one variable (Cliff & Krus, 1976; Darlington, Weinberg, & Walberg, 1973; Thompson, 1980, 1984). For each canonical coefficient, standardized canonical function coefficients and structure coefficients were computed. These coefficients served as *inferential-based effect sizes* (Onwuegbuzie, 2001).

The fifth and final stage of the mixed-methodological analysis involved narrative profile formation. Specifically, the number of average profiles (Tashakkori & Teddlie, 1998) was determined using an ipsative approach in which participants' responses to each theme were interpreted relative to their responses to the other themes (Allport, 1937, 1962, 1966; Block, 1957; Stephenson, 1953) in the following manner: (a) for each participant, the emergent theme scores (i.e., 0 or 1) were ranked such that each scale took on a value from one through six; and (b) the measure of similarity used for the analysis was based on the theme scores ranked from lowest to highest within each profile. An intra-individual correlation matrix was then formed by correlating each pair of profiles, yielding $(n)(n-1)/2$ Spearman Rho values (where n was the number of respondents). This correlation matrix was cluster-analyzed in order that individualistic patterns could be characterized for each preservice teacher. The formation of average profiles represented the *qualitizing* of previously-quantitized themes (Tashakkori & Teddlie, 1998).

Results

Stage 1 and Stage 2 Analyses

The participants listed a total of 125 unique characteristics of effective teachers. Table 1 presents the six themes that emerged from the students' responses: student-centeredness, enthusiasm for teaching, ethicalness, classroom and behavior management, teaching methodology, and knowledge of subject. The overall inter-rater reliability between the two coders pertaining to the categorization of the units into the six themes was .984.

The prevalence rates of each theme (i.e., (manifest) frequency effect sizes, Onwuegbuzie, 2001) also are presented in Table 1. Interestingly, student-centeredness was the most endorsed theme, with nearly 80% of the sample citing one or more traits that fell into this category. Examples of student-centeredness included "love of students," "sensitive," "supportive," "kind," "caring," and "patient"; descriptors of enthusiasm for teaching are "love of subject," "commitment," "untiring," and "true love of job"; examples of ethicalness include "impartial," "unbiased," "honesty," and "fair"; words that describe classroom and behavior management are "authoritative," "good disciplinarian," "observant," and "leadership"; examples that characterize teaching methodology are "knowing how to teach," "variety of teaching methods," and "prompt feedback"; finally, knowledge of subject includes descriptors such as "intelligent," "knowledge," and "smart."

Table 1
Themes Emerging from Preservice Teachers' Perceptions of the Characteristics of Effective Teachers

Theme	Endorsement Rate (%)
Student-centeredness	79.5
Enthusiasm for teaching	40.2
Ethicalness	38.8
Classroom and behavior management	33.3
Teaching methodology	32.4
Knowledge of Subject	31.5

Table 2 presents the number of characteristics cited (i.e., units) for each theme, alongside the proportion of characteristics identified per theme. The proportions noted in Table 2 represented (manifest) intensity effect sizes (Onwuegbuzie, 2001; Onwuegbuzie & Teddlie, in press). It can be seen from Table 2 that student-centeredness contained the most number of characteristics (29.6%), followed by classroom and behavior management (20.0%). The knowledge of subject theme contained the least number of characteristics (4.8%).

Table 3 presents the intercorrelations among the six themes. Interestingly, after applying the Bonferroni adjustment (Onwuegbuzie & Daniel, in press-a), only the correlation between responses to the classroom and behavior management theme and the enthusiasm for teaching theme were statistically significant. Using Cohen's (1988) criteria, this relationship (i.e., $r = .20$, $p < .003$) was small to moderate. Specifically, students who endorsed characteristics classified as being indicative of classroom and behavior management were less likely to cite characteristics representing enthusiasm for teaching.

Table 2
Number and Proportion of Preservice Teachers' Characteristics Cited for Each Theme

Theme	Number of Characteristics	%
Student-centeredness	37	29.6
Enthusiasm for teaching	19	15.2
Ethicalness	16	12.8
Classroom and behavior management	25	20.0
Teaching methodology	22	17.6
Knowledge of Subject	6	4.8
Total	125	

Table 3
Intercorrelations Among all Themes

Variable	1	2	3	4	5
1. Student-centeredness					
2. Ethicalness	-.11				
3. Knowledge of subjects	-.14	-.10			
4. Teaching methodology	-.06	-.03	-.05		
5. Classroom and behavior management	-.05	-.11	-.04	-.08	
6. Enthusiasm for teaching	-.02	-.08	-.07	.03	-.20*

* statistically significant after the Bonferroni adjustment

Stage 3 Analysis

An exploratory factor analysis was used to determine the number of factors underlying the six themes. Specifically, a maximum likelihood factor analysis was used. This technique, which gives better estimates than does principal factor analysis (Bickel & Doksum, 1977), is perhaps the most commonly used method of common factor analysis (Lawley & Maxwell, 1971). As recommended by Kieffer (1999) and Onwuegbuzie and Daniel (in press-b), the correlation matrix in Table 3 was used to undertake the factor analysis. An orthogonal (i.e., varimax) rotation was used because of the low degree of correlation among the themes (cf. Table 3). This analysis was used to extract the latent constructs. As conceptualized by Onwuegbuzie (2001), these factors represented *meta-themes*.

The eigenvalue-greater-than-one rule, also known as K1 (Kaiser, 1958), was implemented to ascertain an appropriate number of factors to retain. This technique resulted in a four factors (i.e., meta-themes). The "scree" test (Cattell, 1966; Zwick & Velicer, 1986) also suggested that four factors be retained. This four-factor solution is presented in Table 4. Using a cutoff correlation of 0.3, recommended by Lambert and Durand (1975), as an acceptable minimum loading value, it can be seen from this table that the following themes loaded significantly on the first factor: classroom and behavior management and enthusiasm; the following themes loaded on the second factor: knowledge of subject and student-centeredness; the following theme loaded on the third factor: ethicalness; and the following theme loaded on the fourth factor: teaching methodology. Clearly, the first meta-theme (i.e., Factor 1) can be labeled *classroom atmosphere*. The second meta-theme can be termed *knowledge of subject and student*. The third meta-theme represents *ethicalness*. Finally, the fourth meta-theme denotes *teaching methodology*. Interestingly, within the *classroom atmosphere* meta-theme (i.e., Factor 1), the enthusiasm for teaching theme was negatively related to the classroom and behavior management theme. Also, within the *subject and student* meta-theme (i.e., Factor 2), the knowledge of subject theme and the student-centeredness themes were negatively related. The thematic structure is presented in Figure 1. This figure illustrates the relationships among the themes and meta-themes arising from preservice teachers' perceptions of the characteristics of effective teachers.

Table 4
Summary of Themes and Factor Loadings from Maximum Likelihood Oblique Factor Analysis: Four-Factor Solution

Theme	Factor Loading[1] 1	2	3	4	Communality Coefficient
Enthusiasm for teaching	**.70**	-.01	.29	-.12	.60
Classroom and behavior management	**-.73**	-.23	.00	.32	.69
Knowledge of Subject	-.28	**.57**	.53	-.30	.77
Student-centeredness	.12	**-.79**	.04	-.20	.69
Ethicalness	.10	.33	**-.85**	-.18	.87
Teaching methodology	.33	.17	.07	**.85**	.87
Trace	1.24	1.14	1.10	1.01	4.49
% of variance explained	20.65	19.07	18.26	16.74	74.72

[1] Coefficients in bold represent loadings with the largest effect size within each theme, using a cut-off loading of 0.3 recommended by Lambert and Durand (1975).

An examination of the *trace* (i.e., the proportion of variance explained, or eigenvalue, after rotation; Hetzel, 1996) revealed that the *classroom atmosphere* meta-theme (i.e., Factor 1) explained 20.65% of the total variance, the *subject and student* meta-theme (i.e., Factor 2) accounted for 19.07% of the variance, the *ethicalness* meta-theme (i.e., Factor 3) explained 18.26% of the variance, and the *teaching methodology* meta-theme (i.e., Factor 4) accounted for 16.74% of the variance. These four meta-themes combined explained 74.7% of the total variance. This total proportion of variance represents a *latent effect size*, which can be considered very large. The manifest effect sizes associated with the four meta-themes (i.e., proportion of characteristics identified per meta-themes) were as follows: *classroom atmosphere* (64.8%), *subject and student* (88.6%), *ethicalness* (38.8%), and *teaching methodology* (32.4%).

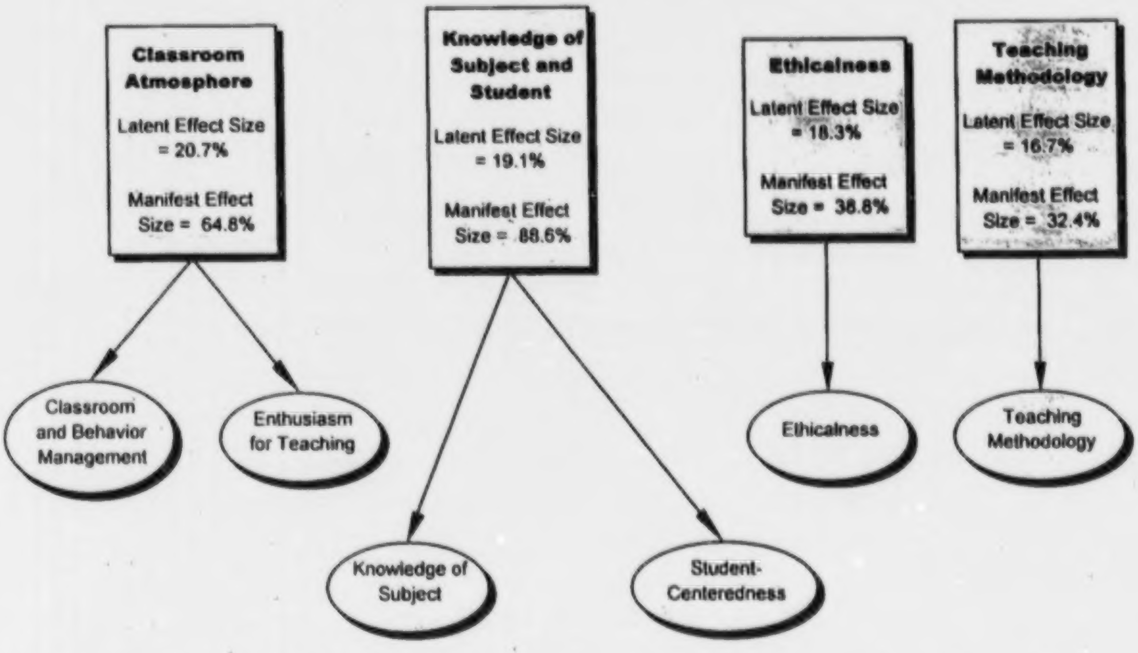

Classroom Atmosphere
Latent Effect Size = 20.7%
Manifest Effect Size = 64.8%
Knowledge of Subject and Student
Latent Effect Size = 19.1%
Manifest Effect Size = 88.6%
Ethicalness
Latent Effect Size = 18.3%
Manifest Effect Size = 38.8%
Teaching Methodology
Latent Effect Size = 16.7%
Manifest Effect Size = 32.4%
Classroom and Behavior Management
Enthusiasm for Teaching
Ethicalness
Teaching Methodology
Knowledge of Subject
Student-Centeredness

Figure 1. Thematic structure pertaining to preservice teachers' perceptions of the characteristics of effective teachers.

Stage 4 Analysis

A series of Fisher's Exact tests, using the Bonferroni adjustment to control for Type I error ($p < .05$), indicated that females tended to place more weight on student-centeredness as a measure of teacher effectiveness than did males, whereas more males than did females tended to endorse management style. Also, older students tended to cite more frequently attributes related to ethicality. Finally, Caucasian-American students tended to endorse management skills more than did minority students.

A canonical correlation analysis was undertaken to examine the relationship between the six themes and a selection of demographic variables. The six themes were treated as the multivariate set of variables, whereas the following variables were utilized as the dependent multivariate profile: gender, age, year of study, ethnicity, type of high school attended by respondent, location of high school attended, and whether the respondent was a parent of a school-aged child.

The number of canonical functions (i.e., factors) that can be generated for a given dataset is equal to the number of variables in the smaller of the two variable sets. Because six themes were correlated with seven independent variables, six canonical functions were generated.

The canonical analysis revealed that the six canonical correlations combined were statistically significant ($p < .05$). However, when the first canonical root was excluded, the remaining five canonical roots were not statistically significant. Similarly, when the first and second canonical roots were excluded, the remaining canonical roots were not statistically significant. Indeed, removal of subsequent canonical roots did not lead to statistical significance. Together, these results suggest that the first canonical function was statistically significant, but the remaining five roots were not statistically significant. However, because the calculated probabilities are sensitive to sample size, particular attention should be paid to the educational (practical) significance of the obtained results (Thompson, 1980). The educational significance of canonical correlations typically are assessed by examining their size (Thompson, 1980, 1984, 1988, 1990). The canonical correlation indicates how much variance the sets of weighted original variables share with each other (Thompson, 1988). In the present study, the first canonical correlation ($R_{c1} = .44$) appeared to be moderately educationally significant, contributing 19.4% (i.e., R_{c1}^2) to the shared variance. Consequently, only the first canonical correlation was interpreted.

Data pertaining to the first canonical root are presented in Table 5. This table provides both standardized function coefficients and structure coefficients. Using a cutoff correlation of 0.3 (Lambert & Durand, 1975), the standardized canonical function coefficients revealed that ethicalness, knowledge of subject, teaching methodology, and classroom and behavior management made important contributions to the set of themes--with classroom and behavior management being the major contributor. With respect to the demographic set, gender, age, year of study, and ethnicity made noteworthy contributions.

Table 5
Canonical Solution for First Function: Relationship Between Six Themes and Selected Demographic Variables

Variable	Standardized Coefficient	Structure Coefficient	Structure^2
Theme:			
Student-centeredness	.146	.091	.008
Ethicalness	.581*	.558*	.311
Knowledge of subjects	-.300*	-.309*	.095
Teaching methodology	.345*	.321*	.103
Classroom and behavior management	-.648*	-.694*	.482
Enthusiasm for teaching	-.173	-.063	.004
Demographic Variable:			
Gender	.423*	.603*	.364
Age	.309*	.104	.011
Year of study	-.648*	-.457*	.209
Ethnicity	-.592*	.514*	.264
Type of high school attended	.155	.231	.053
Area of high school attended	.293	.220	.048
Parental status	-.114	.112	.013

* loadings with effect sizes larger than .3 (Lambert & Durand, 1975)

The structure coefficients revealed that ethicalness, knowledge of subject, teaching methodology, and classroom and behavior management made important contributions (i.e., were practically significant) to the first canonical variate. The square of the structure coefficient indicated that these variables explained 31.1%, 9.5%, 10.3%, and 48.2% of the variance, respectively. With regard to the demographic cluster, gender made the strongest contribution, with ethnicity and year of study making moderate contributions. The square of the structure coefficient indicated that gender, ethnicity, and year of study explained 36.4%, 26.4%, and 20.9% of the variance, respectively.

According to Thompson (in press), variables with small structure coefficients, but standardized coefficients that are large in absolute value magnitude indicate that they are *suppressor* variables in the canonical correlation model. Suppressor variables are variables that assist in the prediction of dependent variables due to their correlation with other independent variables (Tabachnick & Fidell, 1996). In the present study, age appeared to serve as a suppressor variable because the standardized coefficient associated with this variable was large, whereas the corresponding structure coefficient for age was relatively small. It is likely that age was a suppressor variable because of its relationship with one or more of the other demographic variables. In particular, age had an extremely large correlation with the respondent's parental status ($r = .78, p < .0001$), with older students having a greater tendency to be a parent. Thus, age improved the predictive power of demographic variables by suppressing variance that was irrelevant to this prediction as a result of its relationship with these two variables.

In sum, the results of the canonical correlation analysis suggest that females, college-level juniors, and minority students tended to endorse teacher characteristics that were classified as ethical and teaching methodology to a greater extent than did their counterparts. They also tended to rate attributes that were associated with knowledge of subject and classroom and behavior management to a lesser degree. Age served as a suppressor variable.

Finally, the quantitized dichotomous variables that formed the six themes were qualitized via narrative profile formation (Tashakkori & Teddlie, 1998). Specifically, the number of average profiles was determined using an ipsative approach in which participants' respones to each theme were interpreted relative to their responses to the other themes (Allport, 1937, 1962, 1966; Block, 1957; Stephenson, 1953) in the following manner: (a) for each participant, the six theme scores (i.e., 0 or 1) were ranked such that each scale took on a value from one through six; and (b) the measure of similarity used for the analysis was based on the theme scores ranked from lowest to highest within each profile.

An intra-individual correlation matrix was then formed by correlating each pair of profiles, yielding 23,871 (i.e., 219 x 218/2) Spearman Rho values. This correlation matrix was cluster-analyzed utilizing the VARCLUS procedure of the Statistical Analysis System (SAS Institute Inc., 1990) in order that individualistic patterns could be characterized for each preservice teacher. Respondents having similar profiles were expected to cluster together. The criterion of percentage variation explained by each cluster decided the most meaningful cluster solution.

In an attempt to obtain the minimum cluster solution that explained the maximum variation, the criterion of terminating the splitting of clusters when each cluster has only one eigenvalue greater than one was applied. Also, cluster solutions that added less than 5% to the explained variation were eliminated from consideration. Thus, a four-cluster solution, which explained 63.5% of the variation, was selected as the most meaningful and parsimonious.

The profiles for the resulting four clusters are displayed pictorially in Figure 2. The six themes are presented on the horizontal axis, whereas the proportion of students who endorsed one or more teacher characteristics belonging to each theme are presented on the vertical axis. As such, each of the four emergent profiles represented an average set of responses across each theme. As can be seen, members of Cluster 1 (n = 56) were extremely likely to endorse the student-centeredness (probability (p) = .84) and enthusiasm for teaching (p = .71) themes. These preservice teachers were moderately likely to endorse the teaching methodology theme (p = .41). However, they were unlikely to endorse the knowledge of subject (p = .30), classroom and behavior management (p = .16), and ethicalness (p = .11) themes.

Individuals in Cluster 2 (n = 51) also highly rated student-centeredness (p = .83). Additionally, they were very likely to endorse classroom and behavior management (p = .16). However, they were unlikely to cite a characteristic associated with the teaching methodology (p = .27), enthusiasm for teaching (p = .21), ethicalness (p = .18), and knowledge of subject (p = .14) themes. Members of Cluster 3, like the former clusters, highly rated student-centeredness (p = .83). This group also highly rated ethicalness (p = .85). On the other hand, Cluster 3 respondents were unlikely to cite a characteristic associated with the enthusiasm for teaching (p = .37), teaching methodology (p = .25), classroom and behavior management (p = .25), and knowledge of subject (p = .22) themes. Finally, preservice teachers in Cluster 4 were highly likely to endorse the student-centeredness theme (p = .74), although not as uniformly as did members of the other three clusters. They were also likely to cite a teacher characteristic belonging to the knowledge of subject theme (p = .68). They were moderately likely to endorse the ethicalness (p = .40) and teaching methodology (p = .40) themes. However, they were unlikely to endorse the enthusiasm for teaching (p = .32) and classroom and behavior management (p = .30) themes.

Discussion

The purpose of the present study was to determine preservice teachers' perceptions about the characteristics of effective teachers, as well as to investigate factors that may have influenced their responses. Using qualitative and quantitative analytical techniques, the perceptions held by preservice teachers were found to represent a multidimensional construct. Specifically, perceptions were identified that led to the following six themes: student-centeredness, enthusiasm for teaching, ethicalness, classroom and behavior management, teaching methodology, and knowledge of subject.

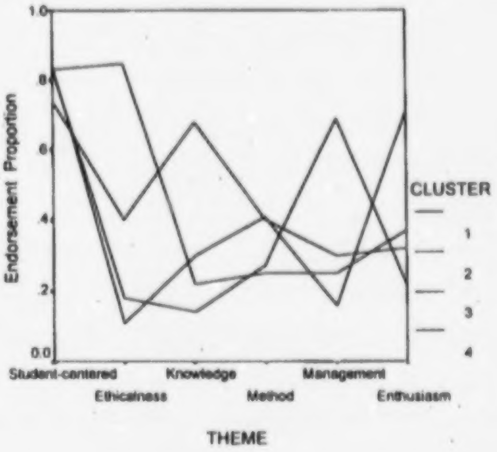


Figure 2. Average profiles relating to preservice teachers' perceptions of the characteristics of effective teachers.

Student-centeredness represented descriptors that received the greatest endorsement. Specifically, nearly 80% of students noted one or more characteristics representing this theme. This proportion, which represents a very large effect size, suggests that the current sample, in general, rated student-centeredness as being the most common characteristic of effective teachers. The remaining five themes were endorsed by a similar proportion of students--between 30% and 40% of the preservice teachers--representing moderate effect sizes. Interestingly, the fact that only 1 of the 15 relationships among the six themes was statistically significant suggests that these themes were somewhat independent of one another.

The finding that student-centeredness was the most prevalent theme endorsed by the preservice teachers provides incremental validity to Wubbels and Brekelmans' (1997) contention that effective teachers are strong in student-teacher relationships and that a solid student-teacher relationship is central for a positive classroom climate. Consistent with Wubbels and Brekelmans, the preservice teachers cited characteristics such as caring, empathetic, respectful, observant, sensitive, and supportive, which are hallmarks of positive student-teacher relationships.

The six themes that emerged from the current data dealt with teacher characteristics (i.e., student-centeredness and enthusiasm for teaching), or were either content-based (i.e., ethicalness and knowledge of subject) or process-related (i.e., teaching methodology and classroom and behavior management). This provides further evidence that the perceptions of teachers represent a complex phenomenon. Moreover, through exploratory factor analysis, the six themes were subdivided into the following four meta-themes: classroom atmosphere, subject and student, ethicalness, and teaching methodology. Both the latent and manifest effect sizes associated with these meta-themes were moderate to large.

Although there is an abundance of literature providing information regarding teacher effectiveness, much of this material consists of long lists of teacher characteristics (e.g., Berliner, 1985; Cotton, 1995; Finn, 1993; Good & Brophy, 1994; Redfield & Rousseau, 1981; Rosenshine & Stevens, 1986; Tobin, 1987; Wortruba & Wright, 1975). Albeit informative, these lists have limited utility because their length somewhat prohibits comparisons from one inventory to the next. Thus, a major contribution of the present investigation is to provide a typology of teacher characteristics. Indeed, this appears to be the first organized framework pertaining to preservice teachers' perceptions of effective teacher traits.

Of the few frameworks presently in existence in the area of teacher characteristics, the themes that emerged from the current inquiry are most consistent with the AASA's two-element conceptualization of effective teachers, namely, (a) management and instructional techniques and (b) personal characteristics (Demmon-Berger, 1986). Specifically, the first three themes in Table 1 (i.e., student-centeredness, enthusiasm for teaching, and ethicalness) reflect the category of personal characteristics, whereas the last three categories (i.e., classroom and behavior management, teaching methodology, and knowledge of subject) can be classified as management and instructional techniques. Comparing the results of the current study to the AASA's conceptualization revealed that a higher proportion (96.8%) of the present sample of preservice teachers noted one or more characteristics representing the personal characteristic domain than rated a trait representing management and instructional techniques (54.8%). Furthermore, McNemar's test indicated no relationship ($Q_m = 2.00; p > .05$) between AASA's two response categories. That is, preservice teachers who rated a personal characteristic as being evidence of an effective teacher were not more likely to rate a management and instructional technique. This suggests that personal characteristics and management and instructional techniques are deemed by preservice teachers to be independent constructs.

Similarly, the present themes can be classified using Roueche et al.'s (1986) three dimensions of interpersonal skills, motivation, and cognitive skills. Specifically, three themes in Table 1 (i.e., student-centeredness, ethicalness, and classroom and behavior management) can be categorized as interpersonal, two themes (i.e., enthusiasm for teaching and teaching methodology) can be classified as motivation, and one theme (i.e., knowledge of subject) can be categorized as cognitive. Consequently, the responses of 94.5% of the present sample can be coded as interpersonal, 58.9% as motivation, and 31.5% as cognitive. McNemar's test indicated no relationship between the response categories of interpersonal and motivation ($Q_m = 0.01; p > .05$), interpersonal and cognitive ($Q_m = 0.60; p > .05$), and cognitive and motivation ($Q_m = 1.16; p > .05$). That is, preservice teachers who rated any one of the three dimensions were not more likely to rate a characteristic falling into either of the remaining two components. This suggests that Roueche et al.'s (1986) three dimensions are deemed by preservice teachers to be independent constructs.

Additionally, the current themes can be categorized using Reed and Bergemann's (1992) two-element conceptualization that effective teacher characteristics can be classified as "measurable" and "not measurable." The knowledge of subject, teaching methodology, and classroom and behavior management themes would be designated as effective measurable traits, whereas student-centeredness, enthusiasm for teaching, and ethicalness would be labeled as not measurable.

A comparison of responses falling into AASA's two-dimensional conceptualization and those falling into Roueche et al.'s (1986) three-component framework and Reed and Bergemann's (1992) two-component representation indicate that preservice teachers, in general, regard the *interpersonal context* as the most important aspect of teaching. Interestingly, this finding is consistent with Onwuegbuzie (1999), who found that of the intelligences identified by Gardner (1983), inservice teachers tended to be most oriented toward interpersonal intelligence.

Fisher's exact tests indicated that females tended to place more weight on student-centeredness as a measure of teacher effectiveness than did males, whereas more males than did females tended to endorse management style. In addition, older students tended to cite more frequently attributes related to ethicality, and Caucasian-American students tended to endorse management skills more than did minority students. These findings, coupled with the result that gender, age, ethnicity, and year of study were related to the teacher characteristics of ethicalness, teaching methodology, knowledge of subject, and classroom and behavior management, indicate the importance of not assuming that all preservice teachers have the same perceptions about what makes a teacher effective.

Using ipsative/cluster analyses, four profiles of students' responses to the six themes emerged. The first profile comprised preservice teachers who believed that student-centeredness and enthusiasm for teaching were central characteristics of effective teachers. For these teachers, the remaining themes received only moderate or low endorsement. The second profile consisted of individuals who highly rated student-centeredness and classroom and behavior management, but who were unlikely to cite a characteristic associated with the remaining four themes. Members of the third profile highly rated student-centeredness and ethicalness, but were unlikely to cite a characteristic associated with the remaining four categories. Finally, preservice teachers belonging to the fourth profile were highly likely to endorse the student-centeredness and knowledge of subject themes. However, they were much less likely to endorse the ethicalness, teaching methodology, enthusiasm for teaching, and classroom and behavior management themes. Teacher educators might consider eliciting such profile information from their preservice teachers and comparing their profiles to their actual teaching styles while student teaching. As such, teacher educators can assist student teachers to reflect on best practices that optimize the preservice teachers' perceptions on characteristics of effective teaching.

Two limitations of the current study must be noted. First, the sample represented preservice teachers attending a college in a geographically-restricted region. Thus, the extent to which these findings generalize to students from other geographic regions is not clear, suggesting a need for replication using more diverse samples. Second, despite the complexity of the statistical analyses used, it should be acknowledged that it does not allow assessment of causality. Consequently, the causal role of the antecedents of student responses identified in the present investigation remains to be established. Future research also might investigate how stable preservice teachers' perceptions are over time. Additionally, investigations are needed comparing perceptions of preservice teachers regarding characteristics of effective teachers to those of teacher trainers, inservice teachers, administrators, and school-age children and parents themselves. Such studies should help to increase our understanding of the similarity of perceptions held by different parties involved in the public school educational process.

The present findings make an important contribution to the effective teacher literature by simultaneously quantifying and qualifying preservice teachers' perceptions of effective teachers. Using mixed methodological data analysis techniques allowed not only the identification of students' perceptions of effective teachers, but also facilitated the computation of effect sizes associated with these perceptions, the determination of the structural relationships among perception categories, and the identification of antecedent correlates of these responses. Thus, future research in this area should continue using this *pragmatist* paradigmatic approach.

References

Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Holt.

Allport, G.W. (1962). The general and the unique in psychological science. *Journal of Personality, 30*, 405-422.

Allport, G.W. (1966). Traits revisited. *American Psychologist, 21*, 1-10.

Berliner, D.C. (1985). Effective classroom teaching: The necessary but not sufficient condition for developing exemplary schools. In G.R. Austin & H. Garber (Eds.), *Research on exemplary schools* (pp. 127-151). Orlando, FL: Academic Press.

Bickel, P.J., & Doksum, K.A. (1977). *Mathematical statistics*. San Francisco: Holden-Day.

Block, J. (1957). A comparison between ipsative and normative ratings of personality. *Journal of Abnormal and Social Psychology, 54*, 50-54.

Caracelli, V.W., & Greene, J.C. (1993). Data analysis strategies for mixed-methods evaluation designs. *Educational Evaluation and Policy Analysis, 15*(2), 195-207.

Cattell, R.B.(1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Cliff, N., & Krus, D.J. (1976). Interpretation of canonical analyses: Rotated vs. unrotated solutions. *Psychometrica, 41*, 35-42.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: John Wiley.

Colaizzi, P.F. (1978). Psychological research as the phenomenologist views it. In R. Vaile & M. King (Eds.), *Existential phenomenological alternatives for psychology* (pp. 48-71). New York: Oxford University Press.

Constas, M.A. (1992). Qualitative data analysis as a public event: The documentation of category development procedures. *American Educational Research Journal, 29*, 253-266.

Cotton, K. (1995). *Effective student practices: A research synthesis--1995 update*. [On-line]. Available http://www.nwrel.org/scpd/esp/esp95.html#1

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publications.

Darlington, R. B., Weinberg, S. L., & Walberg, H. J. (1973). Canonical variate analysis and related techniques. *Review of Educational Research, 42,* 131-143.

Demmon-Berger, D. (1986). *Effective teaching: Observations from research*. Arlington, VA: American Association of School Administrators. (ERIC Document Reproduction Service No. ED 274 087)

Educational Testing Services (1997). Pathwise: A framework for teaching. Retrieved December 1, 1999 from the World Wide Web: http://www.teachingandlearning.org

Finn, J. D. (1993). *School engagement and students at risk*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Goetz, J. P., & Lecompte, M. D. (1984). *Ethnography and the qualitative design in educational research.* New York: Academic Press.

Good, T. L., & Brophy, J. E. (1994). *Looking in classrooms* (6th ed.). New York: HarperCollins.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11,* 255-274.

Hetzel, R. D. (1996). A primer on factor analysis with comments on patterns of practice and reporting. In B. Thompson (Ed.), *Advances in social science methodology* (Vol 4, pp. 175-206). Greenwich, CT: JAI Press.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psyhcometrika, 23,* 187-200.

Kieffer, K. M. (1999). An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. *Research in the Schools, 6*(2), 75-92.

Lambert, Z. V. & Durand, R. M. (1975). Some precautions in using canonical analysis. *Journal of Market Research, XII,* 468-475.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method.* New York: Macmillan Publishing Co., Inc.

Lincoln, Y. S., & Guba, E.G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage.

Moustakas, C. (1994). *Phenomenological research methods*. Thousands Oaks, CA: Sage.

National Board for Professional Teaching Standards. (1994). *What teachers should know and be able to do.* Detroit, MI: Author.

Norton, J. L. (1997, November). *Learning from first year teachers: Characteristics of the effective practitioner*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, TN. (ERIC Document Reproduction Service No. ED 418 050)

Onwuegbuzie, A. J. (1999). *Multiple intelligence profiles of inservice teachers.* Unpublished manuscript, Valdosta State University.

Onwuegbuzie, A. J. (2001, April). *Effect sizes in qualitative research: A prolegomenon.* Paper presented at the annual conference of the American Educational Research Association (AERA), Seattle, WA.

Onwuegbuzie, A. J., & Daniel, L. G. (in press-a). Uses and misuses of the correlation coefficient. *RESEARCH IN THE SCHOOLS, 9*(1).

Onwuegbuzie, A. J., & Daniel, L. G. (in press-b). Typology of analytical and interpretational errors in quantitative and qualitative educational research. *Current Issues in Education.*

Onwuegbuzie, A. J., & Teddlie, C. (in press). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.

Redfield, D., & Rousseau, E. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research, 18*(2), 237-245.

Reed, A. J., & Bergemann, V. E. (1992). *In the classroom: An introduction to education.* Guilford, CT: Dushkin.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376-391). New York: Macmillan.

Roueche, J. E., Baker, G. A., Mullin, P. I., & Boy, N. H. O. (1986). Excellent Teachers. In J.E. Roueche, J.E. & G.A. Baker (Eds.), *Profiling Excellence in Schools* (pp. 87-133). Arlington, VA: American Association of School Administrators.

SAS Institute Inc. (1990). *SAS/STAT User's Guide* (Version 6.12) [Computer software]. Cary, NC: SAS Institute Inc.

Stephenson, W. (1953). *The study of behavior*. Chicago: University of Chicago Press.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper & Row.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Applied Social Research Methods Series (Vol. 46). Thousand Oaks, CA: Sage.

The National Board for Professional Teaching Standards (1987). *Policy position*. [On-Line]. Available http://www.nbpts.org/nbpts/standards/policy.htm.

Thompson, B. (1980, April). *Canonical correlation: Recent extensions for modeling educational processes*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretations*. Newbury Park, CA: Sage Publications. (ERIC Document Reproduction Service No. ED 199 269)

Thompson, B. (1988, April). *Canonical correlation analysis: An explanation with comments on correct practice*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B. (1990, April). *Variable importance in multiple regression and canonical correlation*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 317 615)

Thompson, B. (in press). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding multivariate statistics* (Vol. 2). Washington, DC: American Psychological Association.

Tobin, K. (1987). The role of wait-time is higher cognitive learning. *Review of Educational Research, 24*(1), 69-95.

Wortruba, T. R., & Wright, P. L. (1975). How to develop a teacher-rating instrument: A research approach. *Journal of Higher Education*, 46, 653-663.

Wubbels, T., Levy, J., & Brekelmans, M. (1997 April). Paying attention to relationships. Educational Leadership, *54*(7), 82-86.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

Note

We would like to express our gratitude to Abbas Tashakkori for his extremely helpful comments on an earlier version of this article.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 2, 59-70

Teachers' Beliefs About Mathematics Reform: Instructional Implications for Students With Learning Disabilities

Kathleen M. T. Collins
Saint Mary's University of Minnesota

Michael M. Gerber
University of California Santa Barbara

This study examined the relationship between teachers' beliefs concerning recommendations for effective math instruction/learning, as outlined by the National Council of Teachers of Mathematics (NCTM), and the extent to which teachers believe it is possible to implement the NCTM recommendations into their classroom practices. The study also examined the degree to which (a) teacher personal efficacy and outcome expectancy is influenced by students' self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use); and (b) the degree to which teachers' perceptions of the effectiveness and practicality of grouping strategies, as a measure of classroom practice, are influenced by students' self-regulatory styles. Analyses focused on three groups of matched teachers (n = 39) representing 12 schools. Teachers' beliefs were assessed via the Teachers' Assessment of Mathematics Instruction, a survey instrument developed specifically for this investigation, which utilized Bandura's (1977a, 1977b, 1986) theory of self-efficacy. Nonparametric statistical procedures provided partial support for the hypothesis that self-regulatory styles associated with learning disabilities influence teachers' levels of personal self efficacy and outcome expectancy, as well as their beliefs about the effectiveness and practicality of grouping strategies, despite having beliefs consistent with the NCTM's underlying instructional recommendations. The implications of these findings are discussed.

It is the intent of the National Council of Teachers of Mathematics (NCTM) Standards to assist schools and teachers in transition to what is believed to be superior educational practice by creating a systemic change in how mathematics is viewed and taught by teachers (Commission on Standards for School Mathematics of the NCTM, 1989, 1991, 1995, 1998; see Appendix A). In part, this systematic change in mathematics curricula reflects heightened interest among reformers to foster equity and to promote high standards of academic excellence for "all" students. However, the focus on "mathematics for all" has been critiqued by a number of researchers concerned about the limited reference in the NCTM standards to students with mild disabilities (Giordano, 1993; Hutchinson, 1993; Rivera, 1993). Another criticism is that widespread implementation of the reform curricula has occurred prior to validating its efficacy for students who have learning disabilities (LD) or who are at risk of academic failure in mathematics (Hofmeister, 1993; Kameenui, Chard, & Carnine, 1996; Rivera, 1997).

The intentions and assumptions guiding efforts to restructure mathematics curricula are similar to efforts to create inclusive classrooms. Inclusion is a strong policy movement promoting placement of students with disabilities in general education settings for part or all of the school day. However, research assessing perceptions of regular and special education teachers suggests that teachers generally do not feel prepared to meet the instructional needs of students requiring special instructional accommodations, even though they believe their inclusion is desirable (Semmel, Abernathy, Butera, & Lesar, 1991). Often, opposition to the placement of students with special instructional needs in general education classes has been from successful teachers who intuit that students with special needs require scarce additional resources and are less likely to achieve academically in non-inclusive environments (Gerber & Semmel, 1984). Whinnery, Fuchs, and Fuchs (1991) surveyed 114 general, special, and remedial elementary teachers regarding their perceptions of various instructional and behavioral interventions designed to facilitate inclusion. Their findings revealed that all teachers rated their willingness to implement the intervention significantly higher than their familiarity with or their beliefs regarding ease of actual implementation.

Kathleen M. T. Collins is an assistant professor at Saint Mary's University (Minnesota) and Michael M. Gerber is a professor at the University of California, Santa Barbara. Please address correspondence regarding this article to Kathleen M. T. Collins, Saint Mary's University, 2500 Park Avenue, Minneapolis, MN 55404 or by e-mail to kcollinsknob@cs.com.

Although there was no significant difference in educators' perceptions about the quality or quantity of environmental assistance, teachers rated the likely effectiveness of interventions higher than the ease of implementation.

Moreover, in a recent review of the literature pertaining to the degree that adaptations are implemented in inclusive settings, Scott, Vitale, and Masten (1998) noted inconsistencies between general education teachers' positive perceptions regarding efficacy and feasibility of adapting instruction to learners and the degree to which these teachers utilized individualized instructional techniques when teaching students with LD. Overall, these results indicated that general education teachers favored intact class instruction and were less likely to utilize adapted instructional techniques in general education settings.

However, other research in which teacher beliefs have been investigated in the transition of reform into practice indicates that factors associated with teaching and the instructional environment can both assist and constrain the transition process (Cohen & Ball, 1990; Manouchehri & Goodman, 1998; Putnam, Heaton, Prawat, & Remillard, 1992; Raymond, 1997; Spillane & Zeuli, 1999; Tatto, 1996). For example, several researchers have noted inconsistencies between teachers' perceptions and their observed practices. Specifically, these studies indicated that teachers are willing to accommodate differences in students' learning, but they are less positive concerning their likely efficacy in meeting what they perceive to be a difficult challenge in practice. Other researchers (Schumm & Vaughn, 1991; Schumm, Vaughn, Gordon, & Rothlein, 1994; Whinnery et al., 1991) have reported similar findings regarding teachers' efficacy beliefs in addressing diversity of student learning in practice. Specifically, Schumm and Vaughn (1991) noted inconsistencies in teachers' beliefs about the desirability and feasibility of adapting general education curricula to accommodate students with LD. Their results indicated that teachers perceived all adaptations as more desirable than feasible.

These combined findings suggest that self-efficacy plays a central role in determining teachers' propensity to accommodate agreed-upon curricular policy change while simultaneously accommodating diversity of students' learning styles. Indeed, Social Learning Theory posited by Bandura (1977a, 1977b, 1986) emphasizes the importance of efficacy beliefs in self-regulation of behavior. Self-efficacy, in research on teachers, refers to the structure of beliefs that mediate teachers' selection of actions, degree of effort expended, and the confidence necessary to complete activities that influence student performance (Brophy, 1979). Teachers' self-efficacy beliefs have been conceptualized as a two-dimensional construct defined as personal efficacy of teaching and outcome expectancy of teaching (Ashton & Webb, 1982; Gibson & Dembo, 1984). The first dimension, personal efficacy of teaching, theorizes that teachers with a high level of personal efficacy have confidence in their teaching ability. Teachers with a high level of personal efficacy place a high value on their ability to impact student learning. In contrast, teachers with a low level of personal efficacy place a low value on their ability to impact student learning and experience feelings of helplessness when confronted with obstacles. The second dimension, outcome expectancy of teaching, theorizes that teachers demonstrating a high level of self-efficacy select challenging activities and are self-motivated to pursue their goal regardless of obstacles. In contrast, teachers with a low level of self-efficacy perceive difficult activities as a challenge to their teaching competence. When faced with perceived obstacles, these teachers reduce their level of expended effort and motivation.

Teachers' self-efficacy beliefs are deemed to be a key factor in the transition of policy into practice (Battista, 1994; Little, 1993). For example, self-efficacy beliefs are considered to be a mediating factor in the degree to which teachers utilize instructional innovations (Berman, McLaughlin, Bass, Pauly, & Zellman, 1977). In the domain of mathematics, Thompson (1992) identified other factors, such as the importance of teachers' individual experiences with mathematics and their opportunities to experiment with mathematics, as significant factors influencing teachers' self-efficacy beliefs about teaching and learning. Teachers' self-efficacy beliefs also are considered a mediating factor influencing students' academic outcomes (Bandura, 1977a, 1977b, 1982, 1986, 1989)—particularly students who are perceived as being unmotivated or difficult (Ashton & Webb, 1986). Additionally, teachers' expectations concerning student performance are derived, in part, from individually held beliefs about teaching and learning (Brophy & Good, 1986; Thompson, 1992).

Focus of this Investigation

Restructuring of the mathematics curricula has significant implications for the education of students with LD. The intentions and assumptions guiding efforts to reform mathematics curricula are based on creating meaningful learning opportunities for all students. However, there is limited systematic research to support claims that mathematics curricula and instruction, intended for "all" students (e.g., Standards), translate, in practice, to substantive learning opportunities for students who are at risk of academic failure or who have LD in mathematics.

Similarly, very little is known about the adaptations teachers can or do make to accommodate learning differences in the context of NCTM-guided instruction, and the degree to which this process is influenced by teachers' beliefs concerning effective instruction. Given the national scope of mathematics reform and the degree to which the reform has been criticized for directing minimal attention to the educational needs of students with disabilities, the present investigation sought to provide empirical data concerning teachers' levels of self-efficacy in transitioning reform recommendations into successful instructional opportunities for students with LD. Specifically, the current study was designed to examine the relationship between the extent to which teachers believe that the NCTM (1989) recommendations for math instruction/learning are effective and the extent to which they believe it is possible to implement these recommendations into their classroom practices. Also examined in this current investigation was the degree to which (a) teacher personal efficacy and outcome expectancy, as defined by Bandura's (1977a, 1977b, 1986) theory of self-efficacy, is influenced by self-regulatory styles associated with LD, and (b) the degree to which teachers accommodate students' self-regulatory styles in the context of NCTM guided practice. The following five research questions were posed:

1. What is the relationship between the extent to which teachers believe that the NCTM (1989) recommendations for math instruction/learning are effective and the extent to which they believe it is possible to implement these recommendations into their classroom practices?

2. To what degree is teacher personal efficacy, as measured by their levels of confidence, influenced by students' self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use)?

3. To what degree is teacher outcome expectancy (i.e., expended effort and expectation of student performance) influenced by students' self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use)?

4. To what degree are teachers' perceptions of the effectiveness of grouping strategies influenced by students' self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use)?

5. To what degree are teachers' perceptions of the practicality of grouping strategies influenced by students' self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use)?

This study's hypothesis was guided by Bandura's (1977a, 1977b, 1986) theory of self-efficacy as it relates to teachers' beliefs concerning their efficacy in providing effective instruction. The hypothesis underlying this study was that certain characteristics associated with LD might lead teachers to modify even strongly held beliefs about learning and instruction that underlie NCTM recommendations.

Figure 1 presents a model of teacher self-efficacy as utilized for the present investigation. Important in this model is the conceptualization that personal efficacy, one of the two components of teacher efficacy, comprises teacher confidence, whereas outcome expectancy, the other component of teacher efficacy, consists of expectation of teacher effort and expectation of student performance.

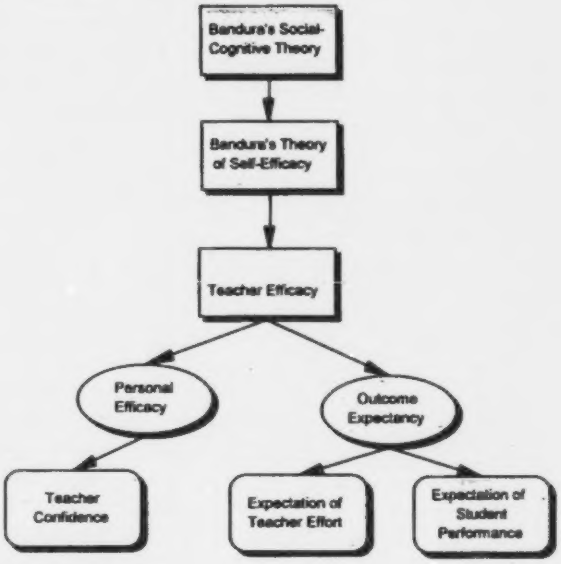

Bandura's Social-Cognitive Theory
Bandura's Theory of Self-Efficacy
Teacher Efficacy
Personal Efficacy
Outcome Expectancy
Teacher Confidence
Expectation of Teacher Effort
Expectation of Student Performance

Figure 1. Model of Teacher Self-Efficacy

Method

Participants

We attempted to control variance associated with measurement of belief by surveying a *matched* sample of teachers ($n = 39$), representing 12 schools. The first sample ($n = 13$), namely, the experimental group (i.e., *inservice project teachers*), consisted of participants in the Tri-County Mathematics Project (TCMP) summer inservice institute. The purpose of the TCMP institute is to assist administrators and teachers in creating curricula that reflect best practices in teaching mathematics. The TCMP provides a supportive environment, via inservice institutes, within which teachers examine and reconstruct their beliefs about mathematics through acquisition of new information, self-reflection, and on-going evaluation. A cohort of inservice project teachers was selected for the present investigation because they represented teachers who were already committed to curricula and instructional reform. It was assumed that such a selection helped to control for teacher motivation and, to some extent, for school-wide resource allocation (e.g., staff inservice training)--critical factors in school-level innovation (Smylie, 1988).

The second sample ($n = 13$) served as one of the two control groups. In obtaining this group of teachers (i.e., *school control group*), inservice project teachers were asked to nominate one colleague within the same school whom they viewed as instructionally compatible (i.e., individuals with whom they would team teach), but who had not had the TCMP summer inservice training. Thus, the major instructional difference between the school control group and the inservice project group was that the former had not attended the recent TCMP inservice training. As such, it was assumed that selection of the school control group of teachers helped to control for school environmental effects with respect to instructional belief systems.

A second control group (i.e., *personality control group*) ($n = 13$) was formed by asking inservice project teachers to nominate another colleague within the same school whom they perceived to be personally, but not necessarily instructionally compatible. The goal of choosing this sample was to minimize personality as a mediating variable influencing teachers' instructional belief systems.

Instrumentation

Teachers' beliefs were assessed via the Teachers' Assessment of Mathematics Instruction (TAMI), which was developed specifically for this inquiry. The TAMI questionnaire consists of four components. The first component comprises 10 stimulus statements pertaining to teaching strategies and instructional goals that were obtained verbatim from reform documents—specifically, the California State Department of Education's (1992) Mathematics Framework and the published NCTM (1989) standards. These statements are presented in Appendix A. Teachers were asked to rate on a 7-point Likert-type scale the extent to which they agreed or disagreed that (a) each of these 10 statements represented effective math instruction and learning (i.e., beliefs about the effectiveness); and (b) it was possible to implement the intent of each statement into their individual classrooms (i.e., beliefs about the practicality).

The second, third, and fourth components of the TAMI each utilize nine instructional vignettes. These vignettes were designed to measure teachers' belief systems with respect to different student self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use). Appendix B presents these nine vignettes. Three vignettes describe students exhibiting poor strategy use while engaged in mathematical problem-solving activities (i.e., poor strategy use). Another three vignettes depict students with poor affect while conducting these activities (i.e., poor affect). The remaining three vignettes characterize students with skills that typified efficient motivation and strategy use (i.e., baseline). Thus, the nine vignettes represented three sets of learning behaviors (each set typified by three vignettes) that teachers recognize as indicative of students' learning styles in contemporary classrooms.

Presentation order of the vignettes was counterbalanced across the three samples of teachers. Development of the vignettes was based on the extant, cognitively-oriented, empirical research literature in the areas of students' self-regulatory characteristics associated with LD related to mathematical problem solving. Additionally, anecdotal behaviors used to contextualize student characteristics were guided, in part, by an adapted version of a survey instrument, the Intervention Strategy Inventory, developed by Larrivee (1985). Asking experienced teachers to review and field-test the questionnaire maximized content-related validity. Modifications to the TAMI were made as necessary.

In the second component, teachers' beliefs were measured by asking teachers to read each vignette and indicate (a) their level of confidence that they felt hypothetical students would reach their instructional goals (i.e., teacher personal efficacy); (b) the amount of instructional effort they were likely to expend (i.e., outcome expectancy); and (c) the degree to which they believed that hypothetical students would reach grade level expectation (i.e., outcome expectancy). Responses to these three indicators were made on a 7-point Likert-type scale, anchored from "very low" to "very high."

With respect to the third component, teachers were asked to indicate the degree to which they believed that seven grouping strategies were an effective response to each of the three sets describing student self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/ strategy use). Grouping strategies were: (a) small group activity involving students of varied ability levels; (b) one-on-one with classmate; (c) one-on-one with teacher or aide; (d) strategy instruction; (e) independent seatwork; (f) intact/whole class instruction; and (g) small group activity involving students of similar ability levels. Responses to the seven instructional options were measured using a Likert-type scale anchored from 1 (not very effective) to 7 (very effective).

In the final component of the TAMI questionnaire, teachers were asked to indicate the degree to which the same seven grouping strategies were a practical response to the three student self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/ strategy use). Responses to these instructional options were measured using a Likert-type scale anchored from 1 (not very practical) to 7 (very practical). For the present investigation, Cronbach coefficient alphas ranged from .88 to .92 for responses across these four components.

Procedure

Graduate researchers telephoned 59 inservice project teachers and invited them to participate in this inquiry. All 59 teachers agreed to participate in the study and thus were mailed a packet containing (a) a cover letter explaining the purpose and importance of their participation in the study and assuring confidentiality regarding responses; (b) support letter from the TCMP coordinators endorsing the research; and (c) three copies of the TAMI questionnaire and three stamped, addressed envelopes to return the completed questionnaires. Inservice project teachers were asked to distribute one questionnaire to each of the two nominated colleagues (i.e., the control group members). To prompt participation, a lottery slip was included, and participants were instructed to return the lottery slip with their completed questionnaire. Lottery prizes for tickets chosen at random were lunch for two and a bookstore gift certificate.

Thus, the sampling frame consisted of 59 inservice project teachers, 59 school control teachers, and 59 personality control teachers. Of the targeted inservice project teachers, 48 (81.4%) returned questionnaires, from which 45 (76.3%) were complete. Of the school control teachers, 42 (71.2%) returned questionnaires, from which 41 (69.5%) were complete. Finally, of the personality control teachers, 17 (28.8%) returned questionnaires, from which 16 (27.1%) were complete. Although the response rates of the inservice project teachers and the school control teachers were high, the relatively low response rate of the personality control teachers led to a total of 13 matches being formed across the three teacher groups. Therefore, the final sample consisted of 39 teachers, with 13 in each group.

A comparison of teachers in these three groups revealed no statistically significant ($p > .05$) differences with respect to educational background (i.e., highest educational degree completed), general teaching experience, special education teaching experience, and grade level they were currently teaching. Specifically, the proportion of teachers whose highest educational attainment consisted of a Bachelor's degree was as follows: inservice project teachers (69.2%), school control teachers (46.2%), and personality control teachers (53.9%). The remaining teachers in each group possessed a MA/M.Ed. degree. With respect to the number of years of general teaching experience, the distribution of teachers who had taught for 10 years or less was as follows: inservice project teachers (38.5%), school control teachers (33.7%), and personality control teachers (33.7%). Further, one teacher in each group reported that they had at least one year of special education teaching experience. Finally, with respect to grade level taught, the majority of teachers in each group taught at the elementary school level: inservice project teachers (76.9%), school control teachers (75.1%), and personality control teachers (92.3%).

Analysis

Because data were not normally distributed, nonparametric statistical procedures were considered most appropriate (Onwuegbuzie & Daniel, in press-a, in press-b). Specifically, analyses consisted of Spearman's rank correlation coefficient, Wilcoxon signed rank test, Hodges-Lehmann test, and the Friedman's two-way analysis of variance test (Marascuilo & McSweeney, 1977). Spearman's rank correlation coefficient and the Wilcoxon signed rank test were used to determine the relationship between teachers' beliefs concerning recommendations for effective math instruction/learning, as outlined by reform documents, and the extent to which they believed it is possible to implement these recommendations into their classroom practices. In the following analyses, teachers' belief systems with respect to different student self-regulatory styles were based on their responses to three sets of vignettes that yielded three composite scores describing the three student self-regulatory styles. Specifically, the composite score pertaining to the first self-regulatory style (i.e., poor strategy use) consisted of the sum of teachers' responses to the three vignettes describing poor strategy use. A second composite score consisted of the sum of teachers' responses to the second set of three vignettes describing poor affect. The final composite score consisted of the sum of teachers' responses to the third set of three vignettes depicting students with skills that typified efficient motivation/strategy use.

The Hodges-Lehmann test was used to determine the relationship between teacher personal efficacy (i.e., levels of confidence) and students' self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use), as well as between teacher outcome expectancy (i.e., expended effort and expectation of student performance) and students' self-regulatory styles. This test also was utilized to determine the relationship between teachers' perceptions of the effectiveness and practicality of grouping strategies and students' self-regulatory styles.

In the Hodges-Lehmann test, teachers' responses per group were aligned on their own group means prior to ranking of responses. As noted by Marascuilo and McSweeney (1977), this statistical procedure avoids confounding the group effects and the treatment effects. Treatment in this study refers to teachers' responses to the three sets of self-regulatory styles portrayed in the vignettes. Tukey pairwise comparisons were computed to interpret significant results arising from the Hodges-Lehmann test ($p < .05$). Corrections were made for tied responses. Finally, a series of Friedman's nonparametric matched sample tests (Marascuilo & McSweeney, 1977) was utilized to determine the relationship between teachers' perceptions of the effectiveness and practicality of grouping strategies and students' self-regulatory styles as a function of teacher group.

Results

Research Question 1: Alignment of Belief and Reform Recommendations

Table 1 presents means and standard deviations pertaining to teachers' beliefs about recommendations for effective math instruction/learning (i.e., beliefs about the effectiveness), and the extent to which they believed it is possible to implement these recommendations into their classroom practices (i.e., beliefs about the practicality of recommendations). Inspection of this table indicates that inservice project teachers were stronger in their professed beliefs concerning both the effectiveness and practicality of reform recommendations. However, a Friedman's two-way analysis of variance test revealed that these beliefs were not statistically significantly higher than that for the remaining two groups. Indeed, no statistically significant difference existed among the teacher groups with respect to either effectiveness ($\chi^2 = 3.89$, $df = 2$, $p > .05$) or practicality ($\chi^2 = 1.08$, $df = 2$, $p > .05$) of reform recommendations.

Table 1
Means, Standard Deviations, and Correlations of Beliefs About Effectiveness With Beliefs About Practicality of Reform Recommendations For Each Teacher Group

	Effectiveness		Practicality		
Teacher Groups	*M*	*SD*	*M*	*SD*	R_s
Inservice Project ($n = 13$)	63.0	4.12	59.1	7.14	.56*
School Control ($n = 13$)	61.4	5.40	58.0	6.80	.85**
Personality Control ($n = 13$)	56.0	10.10	53.0	8.27	.75***

* $p < .05$, ** $p < .01$, *** $p < .001$

To test the equality of sample means with respect to both effectiveness and practicality, pairwise comparisons were conducted using the Wilcoxon signed rank test. Pairwise comparisons revealed that inservice project teachers were statistically significantly more positive about the effectiveness of reform recommendations than were personality control teachers. The effect size associated with the differences in teachers' beliefs about effectiveness was .98, which suggests a large effect (Cohen, 1988). However, no statistically significant difference in level of positiveness about the effectiveness of reform recommendations was found between inservice project teachers and school control teachers, or between personal control teachers and school control teachers. With respect to teachers' beliefs about the practicality of the reform recommendations, no statistically significant difference was found among the three teacher groups.

Wilcoxon's signed rank test also revealed that, across the three groups, teachers reported statistically significantly more positive beliefs pertaining to the effectiveness of the reform recommendations than they did concerning the practicality of these recommendations ($z = 4.11$, $p < .01$). This difference also was noted for the inservice project group ($z = 2.16$, $p < .05$), school control group ($z = 2.81$, $p < .01$), and the personality control group ($z = 2.32$, $p < .05$).

Spearman's rank correlational coefficients indicated, for each teacher group, a statistically significant relationship between teachers' beliefs about effectiveness and beliefs about the practicality of recommendations (Table 1). Using Cohen's (1988) criteria, these coefficients indicate large relationships. Interestingly, the largest association was noted for the school control group, whereas the smallest relationship, albeit large, was observed for the inservice project teachers.

Research Questions 2 and 3: Relationship Between Personal Efficacy and Outcome Expectancy and Students' Self-Regulatory Styles

The Hodges-Lehmann test was used to assess teacher personal efficacy and outcome expectancy with respect to the three self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use) presented via vignettes. The Hodges-Lehmann test indicated a statistically significant difference in teachers' levels of confidence in response to the three self-regulatory styles across the three samples of teachers ($W = 69.15$, $df = 2$, $p < .001$). Because W represents a chi-square value, Cramer's V (i.e., $\sqrt{W/n}$) was utilized as a measure of effect size. Thus, the effect size associated with the differences in levels of confidence was 1.33, which suggests a very large effect (Cohen, 1988). Tukey's post-hoc pairwise comparisons revealed the mean rank associated with poor strategy use (47.00) and poor affect (34.90) was significantly *lower* than the mean rank denoting efficient motivation/strategy use (95.10). However, no statistically significant difference was found between poor strategy use and poor affect self-regulatory styles.

Similarly, the Hodges-Lehmann test indicated a statistically significant difference in teachers' expended effort in response to the three self-regulatory styles across the three samples of teachers ($W = 69.09$, $df = 2$, $p < .001$). The effect size associated with the differences in expended effort was 1.33,which suggests a very large effect (Cohen, 1988). Tukey's post-hoc pairwise comparisons revealed the mean rank associated with poor strategy use (67.44) and poor affect (86.10) was significantly *higher* than the mean rank denoting efficient motivation/strategy use (23.50). There was also a statistically significant difference between poor strategy use and poor affect self-regulatory styles.

Finally, the Hodges-Lehmann test indicated a statistically significant difference in teachers' expectation of student performance in response to the three self-regulatory styles across the three samples of teachers ($W = 58.41$, $df = 2$, $p < .001$). The effect size associated with the differences in expended effort was 1.22, which suggests a very large effect (Cohen, 1988). Tukey's post-hoc pairwise comparisons revealed the mean rank associated with poor strategy use (47.12) and poor affect (37.30) was significantly *lower* than the mean rank denoting efficient motivation/strategy use (92.60). No statistically significant difference was found between poor strategy use and poor affect self-regulatory styles.

Research Questions 4 and 5: Relationship Between Teachers' Perceptions of the Effectiveness and Practicality of Grouping Strategies and Students' Self-Regulatory Styles

Effectiveness of Grouping Strategies. The Hodges-Lehmann test indicated a statistically significant difference in teachers' ratings of seven grouping strategies as an effective response to the three self-regulatory styles across the three samples of teachers ($W = 88.32$, $df = 6$, $p < .001$). Interestingly, the effect size associated with the differences in effectiveness ratings was 1.50, which suggests a very large effect (Cohen, 1988). Tukey's post-hoc pairwise comparisons revealed that the mean rank associated with poor strategy (373.30) and poor affect (338.60) was significantly lower than the mean rank denoting efficient motivation/strategy use (518.10). However, no statistically significant difference was found between poor strategy use and poor affect self-regulatory styles.

Friedman's nonparametric matched-sample test (Marascuilo & McSweeney, 1977) indicated an identical pattern for each teacher group with respect to their ratings of the seven grouping strategies as an effective response to the three self-regulatory styles (Table 2). Specifically, for each teacher group, a statistically significant difference emerged in teachers' ratings of the seven grouping strategies in response to both poor strategy use and poor affect. (All effect sizes were very large.) In contrast, teachers' ratings of grouping strategies did not differ statistically significantly in response to self-regulatory style associated with efficient motivation/strategy use.

Practicality of Grouping Strategies. The Hodges-Lehmann test indicated a statistically significant difference in teachers' ratings of the practicality of the seven grouping strategies in response to the three self-regulatory styles across the three samples of teachers ($W = 35.83$, $df = 6$, $p < .001$). Moreover, the effect size associated with the differences in effectiveness ratings was 0.96, which suggests a very large effect (Cohen, 1988). Tukey's post-hoc pairwise comparisons revealed that the mean rank associated with poor strategy use (393.20) and poor affect (360.00) was significantly lower than the mean rank denoting efficient motivation/strategy use (477.00). However, no statistically significant difference was found between poor strategy use and poor affect self-regulatory styles.

Table 2

Chi-Square Statistics (and Cramer's V Effect Sizes) From Friedman Tests of the Differences in Teachers' Ratings of the Effectiveness of the Seven Grouping Strategies as a Function of Students' Self-Regulatory Styles and Teacher Group

Teacher	Poor Strategy Use	Poor Affect	Baseline
Inservice Project ($n = 13$)	42.11* (1.80)	46.69* (1.90)	7.66
School Control ($n = 13$)	32.47* (1.58)	38.40* (1.72)	11.22
Personality Control ($n = 13$)	33.04* (1.59)	34.29* (1.62)	10.02

$df = 6$; $^*p < .01$.

Interestingly, Friedman's nonparametric matched sample test indicated a differential response pattern for each teacher group with respect to their ratings of the practicality of the seven grouping strategies *vis-a-vis* the three self-regulatory styles (Table 3). Specifically, for each teacher group, a statistically significant difference emerged in teachers' ratings of the seven grouping strategies in response to both poor strategy use and poor affect for inservice project and school control teachers only. In contrast, no statistically significant difference in personality control teachers' ratings of grouping strategies was noted with respect to poor strategy use and poor affect self-regulatory styles. However, across all teacher groups, results indicated a statistically significant difference in response to self-regulatory style associated with efficient motivation/strategy use.

Table 3
Chi-Square Statistics (and Cramer's V Effect Sizes) From Friedman Tests of the Differences in Teachers' Ratings of the Practicality of the Seven Grouping Strategies as a Function of Students' Self-Regulatory Styles and Teacher Group

Teacher	Poor Strategy Use	Poor Affect	Baseline
Inservice Project	27.25*	20.61*	30.50*
(n = 13)	(1.45)	(1.26)	(1.53)
School Control	28.43*	31.37*	39.71*
(n = 13)	(1.48)	(1.55)	(1.75)
Personality Control	4.59	5.55	17.79*
(n = 13)			(1.17)

$df = 6$; $^{*}p < .01$.

Discussion

The first purpose of the present investigation was to determine the relationship between the extent to which teachers believe that the NCTM (1989) recommendations for math instruction/learning are effective and the extent to which they believe it is possible to implement these recommendations into their classroom practices. Results indicated that inservice project teachers' beliefs were strongly consistent with NCTM recommendations, with an overall mean of 6.1 on a 7-point scale. Nevertheless, inservice project teachers were significantly stronger in their professed beliefs concerning the effectiveness of reform recommendations than were the other two teacher groups. Bearing in mind that the only consistent difference between the experimental group and the two control groups was the fact that the former had recently participated in the TCMP institute, it is likely that the summer institute was, at least in part, responsible for the inservice project teachers' relatively more positive attitudes toward the efficacy of the reform recommendations. As such, this particular finding is encouraging.

However, across the three groups, as well as within each group, teachers reported statistically significantly more positive beliefs pertaining to the effectiveness of the reform recommendations than they did concerning the practicality of these recommendations. This finding indicates that, although the teachers believe in the effectiveness of reform recommendations, they have a different set of beliefs as to the degree to which they feel that these recommendations can be implemented. This further suggests that teachers, to some extent, deem the reform recommendations to represent quixotic goals. Moreover, the correlation analyses measuring the alignment of inservice project teachers' beliefs about effectiveness and beliefs about the practicality of recommendations (Table 1) revealed the smallest relationship for the inservice project teachers. This result indicates that the incongruency between effectiveness and practicality beliefs is greatest for inservice project teachers. Unfortunately, it is beyond the scope of the present investigation to determine why the inservice project teachers were relatively less positive about the extent to which the reform recommendations can be incorporated into their classrooms. Thus, this should be the subject of future research.

The apparent difference in beliefs between the effectiveness and practicality of the reform recommendations is consistent with the extant literature. For example, Semmel et al. (1991) found that, even though teachers believe that the inclusion of students who require special instructional accommodations is desirable, they feel unprepared to meet the needs of these children. Similarly, other researchers (e.g., Schumm & Vaughn, 1991; Schumm et al., 1994; Scott et al., 1998; Whinnery et al., 1991) have noted that teachers perceive instructional adaptations for students with LD as more desirable than feasible. Indeed, many of these studies implicate low teacher self-efficacy beliefs as an important reason (i.e., mediating factor) for this discrepancy between desirability and feasibility of adapting general education curricula to accommodate students with LD.

Thus, it is possible that the difference in perceptions noted in the present investigation stems, at least in part, from such low self-efficacy beliefs relating to personal efficacy and outcome expectancy. Yet, this would seem to contradict the fact that the inservice project teachers reported the largest discrepancy. In fact, one would expect the inservice group of educators to have the highest levels of self-efficacy beliefs as a result of participating in the TCMP summer institute, wherein teachers and administrators are helped to create curricula that reflect best practices in teaching mathematics. Why then would the inservice project teachers, presumably with higher levels of self-efficacy beliefs, report the largest discrepancy between effectiveness and practicality beliefs pertaining to the reform recommendations? One reason may be that the recent TCMP inservice experience sensitizes teachers to the various factors (e.g., time, resources, support, student learning styles) that mediate the transition of reform recommendations into practice. In other words, it is likely that the TCMP makes teachers more cognizant of the potential difficulties in implementing effective instruction within their particular classrooms. Inservice project teachers may be more likely to incorporate these mediating factors into their personal efficacy and outcome expectancy beliefs. Thus, future research should investigate the role of these potential mediating factors in the formation of teacher efficacy beliefs as they relate to the successful accommodation of students with LD.

Also examined was the degree to which teacher personal efficacy and outcome expectancy is influenced by students' self-regulatory styles (i.e., poor strategy use vs. poor affect vs. efficient motivation/strategy use). Results indicated that teachers expressed less confidence in their efficacy (i.e., personal efficacy) at addressing students with poor affect and poor strategy use in contrast to baseline students. In addition, teachers perceived that they would have to expend a higher degree of instructional effort (i.e., outcome expectancy) in order for students with poor affect to reach grade level expectation in mathematics. Similarly, teachers had lower expectations of student performance (i.e., outcome expectancy) for poor strategy use and poor affect regulatory style students in contrast to baseline students.

These three sets of findings suggest that student-learning characteristics significantly influence teachers' efficacy beliefs. Consistently, teachers reported relatively low personal efficacy and outcome expectancy when confronted with scenarios in which students exhibited characteristics associated with LD (i.e., poor strategy use and poor affect). A question that is beyond the scope of the present investigation was whether self-regulatory styles affected the teachers' efficacy beliefs consciously or subconsciously. To determine teachers' levels of consciousness in forming their efficacy beliefs, future versions of the TAMI should include open-ended questions asking respondents to provide a reason for each selection.

However, regardless of the level of consciousness, these findings cast serious doubt about the degree to which teachers feel prepared to implement reform recommendations. Interestingly, the fact that the teachers reported lower personal efficacy and outcome expectancy for students who exhibit characteristics associated with LD might explain, in part, why they held less positive beliefs concerning the practicality of the reform recommendations than concerning the effectiveness. The present study also examined the degree to which teachers' perceptions of the effectiveness and practicality of grouping strategies, as a measure of classroom practice, are influenced by students' self-regulatory styles. Responses from inservice project teachers indicated that they recognize self-regulatory learning styles as a significant factor influencing "how" they select grouping strategies in practice. This practical accommodation was consistent with that of school control group teachers. In contrast, however, inservice project teachers' beliefs about applications of grouping strategies were inconsistent with those held by personality control teachers.

In summary, self-regulatory styles associated with LD influence teachers' levels of personal self efficacy and outcome expectancy, as well as their beliefs about the effectiveness and practicality of grouping strategies. These results provide incremental validity to the hypothesis that self-regulatory characteristics associated with LD influence teachers to modify instructional practices, despite beliefs consistent with NCTM's (1989) underlying instructional recommendations. Additionally, these findings support the conclusions of Schumm and Vaughn (1991), Schumm et al. (1994), and Whinnery et al. (1991) regarding teachers' efficacy beliefs in addressing diversity of student learning in practice. These researchers concluded that teachers express support and are amenable to the notion of adapting instruction to address diversity of student learning in the classroom, however, they are less certain concerning their likely efficacy in implementing these adaptations in practice.

The current findings raise questions about the degree that teachers feel prepared to implement instructional techniques aligned to reform recommendations in the context of classroom practice while simultaneously addressing student diversity. Given the national scope of reform-oriented curricula and the likely inclusion of students with LD in general education classes, our findings validate the importance of including guidelines in reform documents (i.e., Standards). These guidelines, which could be emphasized in staff inservice training programs, would provide teachers with explicit strategies for implementing the reform curricula and simultaneously addressing learning styles of students with LD. Hopefully, such specifications would increase teachers' levels of personal efficacy and outcome expectancy in their instruction of students with LD and students at risk of failure in mathematics during the transition period from traditional to reform practice.

A limitation of the of the study stemmed from the fact that some (34.6%) of the control group members had participated in professional development activities that were specifically focused on designing and implementing effective mathematics curricula (i.e., TCMP inservice training). This occurrence threatened both the internal (i.e., history by treatment interaction) and the external validity (i.e., seepage effect) of the findings (Onwuegbuzie, in press). However, these control group teachers participated in the TCMP professional development activities at an earlier time period than did the experimental group teachers. Thus, it is likely that the passage of time differentiated the professional development experience of the control group teachers who participated in the earlier TCMP and the experimental group teachers, thereby minimizing this threat to internal and external validity. An additional threat to internal validity included differential selection of participants. However, as noted above, the experimental and two control groups did not differ with respect to educational background, general teaching experience, special education teaching experience, and grade level they were currently teaching.

Generalizability, specifically population validity and ecological validity, of these results may be limited by the use of a small, non-random sample from a geographically-restricted region. Also, the use of analogs may have affected the external validity of the findings via reactive arrangements (Onwuegbuzie, in press). Therefore, replication of the present study is needed utilizing larger samples of teachers. Future research incorporating mixed methodologies (i.e., quantitative and qualitative research techniques) will provide a broader perspective of the transition of reform curricula into practice for students with LD.

References

Ashton, P. T., & Webb, R. B. (1982, March). *Teachers' sense of efficacy: Toward an ecological model.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York: Longman.

Bandura, A. (1977a). *Social learning theory.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A. (1977b). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122-143.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1989). Regulation of cognitive processes through perceived self efficacy. *Developmental Psychology, 25*, 729-735.

Battista, M. T. (1994, February). Teacher beliefs and the reform in mathematics education. *Phi Delta Kappan*, 462-470.

Berman, P., McLaughlin, M., Bass, G., Pauly, E., & Zellman, G. (1977). *Federal programs supporting educational change, Vol. 7: Factors affecting implementation and continuation* (Report No. R-1589/7-HEW). Santa Monica, CA: The RAND Corporation.

Brophy, J. (1979). Teacher behavior and its effect. *Journal of Educational Psychology, 71*, 733-750.

Brophy, J., & Good., T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook on research on teaching* (pp. 328-375). New York: MacMillan.

California State Department of Education. (1992). *Mathematics curriculum framework for California public schools.* Sacramento, CA: Author.

Cohen, D. K., & Ball, D. L. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis, 12*(3), 347-353.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: John Wiley.

Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*(3), 137-148.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 569-582.

Giordano, G. (1993). Fourth invited response: The NCTM standards: A consideration of the benefits. *Remedial and Special Education, 14*(6), 28-32.

Hofmeister, A. M. (1993). Elitism and reform in special education. *Remedial and Special Education, 14*(6), 8-13.

Hutchinson, N. (1993). Second invited response: Students with disabilities and mathematics educational reform. *Remedial and Special Education, 14*(6), 20-23.

Kameenui, E. J., Chard, D.J., & Carnine, D.W. (1996). Response: The new school mathematics and the age-old dilemma of diversity: Cutting or untying the Gordian knot. In M.C. Pugach & C.L. Warger (Eds), *Curriculum Trends, Special Education, and Reform: Refocusing The Conversation* (pp. 68-83). New York, NY: Teachers College Press.

Larrivee, B. (1985). *Effective teaching for successful mainstreaming.* New York: Longman.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis, 15*(2), 129-151.

Manouchehri, A., & Goodman, T. (1998). Mathematics curriculum reform and teachers: Understanding the connections. *The Journal of Educational Research, 92*, 27-41.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences.* Monterey, CA: Brooks/Cole.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1998). *Principles and standards for school mathematics: Discussion draft.* Reston, VA: Author.

Onwuegbuzie, A.J. (in press). Expanding the framework of internal and external validity in quantitative research. *RESEARCH IN THE SCHOOLS.*

Onwuegbuzie, A.J., & Daniel, L.G. (in press-a). *Uses and misuses of the correlation coefficient. RESEARCH IN THE SCHOOLS.*

Onwuegbuzie, A.J., & Daniel, L.G. (in press-b). Typology of analytical and interpretational errors in educational research. *Current Issues in Education.*

Putnam, R. T., Heaton, R. M., Prawat, R. S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. *The Elementary School Journal, 93*(2), 213-228.

Raymond, A.M. (1997). Inconsistency between teacher beliefs and practices. *Journal for Research in Mathematics Education, 20*, 550-576.

Rivera, D. M. (1993). Third invited response: Examining mathematics reform and the implications for students with mathematical disabilities. *Remedial and Special Education, 14*(6), 24-27.

Rivera, D.M. (1997). Mathematics education and students with learning disabilities: Introduction to the special series. *Journal of Learning Disabilities, 30*(1), 2-19.

Schumm, J.S., & Vaughn, S. (1991). Making adaptations for mainstreamed students: Regular classroom teachers' perspectives. *Remedial and Special Education, 12*(4), 18-27.

Schumm, J.S., Vaughn, S., Gordon, J., & Rothlein, L. (1994). General education teachers beliefs, skills, and practices in planning for mainstreamed students with learning disabilities. *Teacher Education and Special Education, 17*(1), 22-37.

Scott, B.J., Vitale, R., & Masten, W.G. (Mar/Apr. 1998). Implementing instructional adaptations for students with disabilities in inclusive classrooms: A literature review. *Remedial and Special Education, 19*(2), 106-19.

Semmel, M. I., Abernathy, T. V., Butera, G., & Lesar, S. (1991). Teacher perceptions of the Regular Education Initiative. *Exceptional Children, 58*(1), 9-24.

Smylie, M.A. (1988, Spring). The enhancement function of staff development: Organizational and psychological antecedents to individual teacher change. *American Educational Research Journal, 25*, 1-30.

Spillane, J.P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis, 21*(1), 1-28.

Tatto, M. T. (1996). Examining values and beliefs about teaching diverse students: Understanding the challenges for teacher education. *Educational Evaluation and Policy Analysis, 18*(2), 155-180.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: Macmillan.

Whinnery, K. W., Fuchs, L. S., & Fuchs, D. (1991). General, special, and remedial teachers' acceptance of behavioral and instructional strategies for mainstreaming students with mild handicaps. *Remedial and Special Education, 12*(4), 6-13.

Appendix A
Reform Recommendations

Source: *NCTM Standards* (1989)

A demonstration of good reasoning should be rewarded even more than students' ability to find correct answers. (p. 6)

Curriculum should include deliberate attempts, through specific instructional activities, to connect ideas and procedures both among different mathematical topics and with other content areas. (p. 11)

Classroom activities should provide students the opportunity to work both individually and in small and large group arrangements. (p. 67)

Traditional teaching suggests that skill in computation is a precursor to solving problems. Strategies of teaching should be reversed; knowledge often emerges from experience with problems. (p. 10)

In developing problem situations, teachers should emphasize the application of mathematics in real-world problems. (p. 66)

Students should have opportunities to formulate problems and questions that stem from their own interests. (p. 1)

Source: *California Mathematics Framework* (1992)

Real mathematics is seldom done in isolation. If the task is at all difficult or requires serious thought it is best done in collaboration. (p. 16)

Teachers need to organize their classrooms so that students think and talk about their work. (p. 50)

It is not enough for students to produce answers to pre-organized exercises; they must use mathematics to help make sense of real situations. (p. 16)

In an empowering program no student will have to do simplified material. (p. 41)

Appendix B
Instructional Vignettes

Poor strategy use

Inefficient self-monitoring. While solving math problems, Ali devises her own way of solving the problem that indicates that she understands part of the process. However, she has difficulty completing all the steps to a math problem. It is apparent to you that she does not check her work prior to handing it in to you. Ali's family owns a local restaurant. She is learning how to cook. Last week, Ali proudly brought you a box of homemade cookies.

Lack of generalization. In math class May is often inconsistent in the quality of the work that she produces. She demonstrates that she understands how to solve a math problem; but she becomes confused if a problem requiring the same skills is presented in a different format. May enjoys sports and is a member of the city sponsored team.

Short term memory problems. During math class Jeff has problems interpreting your directions and he frequently forgets them. In class he often chooses to ignore directions about how to complete the assignment. Consequently, he usually is the last to complete the assignment in comparison to other students in the class. When Jeff won three checker games in a row during free time, you noticed that he was visibly happy all the rest of the day.

Poor affect

Dependent approach to learning. Sue rarely works without assistance. She frequently approaches you for directions and to check her work. It usually takes her more time in comparison to her peers to complete any task. It is difficult for you to find tasks that interest her. Sue is a member of the local girl scout troop. She is proud of the safety badge that she recently earned.

Low motivation. After being given an assignment, Jon stops trying quickly without further aid and support from you. While seated at his desk he sighs and finds things other than the assignment to occupy his time. You've met Jon's dad and he seemed to be a pleasant person. Jon's mother works part-time at the library. Both parents always listen attentively during conferences.

Negative attitude toward school. Amy frequently protests about the assigned math problems. Often she does not complete her assignments. When asked why the assignment is not finished, she often responds that there are too many problems to answer or that other students have been distracting her during math class. Amy is excited about art and drawing. She often decorates the cover of her notebook with original designs.

Baseline

Appropriate self-monitoring skill. Ken demonstrates that he is organized as far as having the appropriate tools or supplies to complete the task. In approaching a task he devises an appropriate system or series of steps to successfully complete the assignment. He usually checks his work prior to giving it to you. Ken loves to play basketball during recess. Yesterday he told you that his family is going to the Los Angeles Forum to watch the Lakers play basketball.

Independent learner. After being given an assignment, Rob readily begins and completes the work with little assistance from you. Rob can work alone and he can work with other students. He enjoys a challenge and he approaches new tasks with enthusiasm. Rob's family is going on a weekend camping trip. He is excited about the trip. You heard him say that he is in charge of putting up the tent.

Motivated learner. Eva usually completes her assignment on time. She approaches all tasks with a similar level of motivation and interest. Generally, she is quick to understand the key math concept(s) presented in the assignment with a minimum of instructions from you. Eva is learning how to ride a horse. She has riding lessons every Saturday. Yesterday she told you that she wanted a horse for her birthday.

Footnote

This research was supported by the Special Education Policy Research and Exchange Project – Office of Special Education Programs, U.S. Department of Education Grant No. H029D20030. The opinions and content in the paper, however, are those of the authors and do not necessarily reflect the official position of the Department.

Footnote

The authors would like to express gratitude to the teachers that participated in this study.

Footnote

Special thanks to Anthony J. Onwuegbuzie for his helpful comments and recommendations on previous drafts of this manuscript.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 2, 71-81

Academic Success as a Function of the Gender, Class, Age, Study Habits, and Employment of College Students

William J. Lammers
University of Central Arkansas

Anthony J. Onwuegbuzie
Howard University

John R. Slate
University of Texas at El Paso

In this study, college students' predominant study skill strengths and weaknesses were identified. Students' study skills also were examined as a function of age, gender, year in college, academic achievement, study time, and employment status. Participants were 366 undergraduate students enrolled in nine different courses within the College of Education at a university in the mid-South. Students responded positively to only 53.0% of the statements measuring study skills, with study skill weaknesses being identified in the areas of note-taking, reading skills, and time management. Additionally, study skills were related positively to age, GPA, and the number of hours spent studying each week, and related negatively to the number of hours spent working each week. A series of discriminant analyses led to the identification of specific study behaviors that discriminated various subgroups. Implications for intervention programs are discussed.

Many factors relate to academic success in college students. Understanding the relationships among these factors has important implications for intervention programs. Researchers consistently have reported a positive relationship between study habits and academic success (Agnew, Slate, Jones, & Agnew, 1993; Elliot, Godshall, Shrout, & Witty, 1990; Jones, Green, Mahan, & Slate, 1993; Jones, Slate, & Kyle, 1992; Jones et al., 1994; Jones, Slate, & Marini, 1995; Kleijn, van der Ploeg & Topman, 1994). In particular, Jones, Slate, Perez, and Marini (1996), based on a series of studies conducted by Jones and Slate (1992), reported that study skills account for approximately 15% of the variance in undergraduate students' grades. The relationship between study skills and academic achievement has been found at the high school level (Jones, Slate, Bell, & Saddler, 1991; Jones, Slate, Blake, & Holifield, 1992; Slate, Jones, & Dawson, 1993), the undergraduate level (Agnew et al., 1993; Jones et al., 1992; Jones et al., 1994; Jones, Slate, Marini, & DeWater, 1993; Lawler-Prince, Slate, & Jones, 1993; Slate, Jones, & Charlesworth, 1990), and even at the graduate level (Onwuegbuzie, Slate, Paterson, Watson, & Schwartz, 2000). Unfortunately, poor study skills have been noted at both secondary (Jones et al., 1991; Jones et al., 1992; Slate et al., 1993) and post-secondary levels (Onwuegbuzie et al., 2000), with only between 40% and 46% of appropriate study behaviors being performed by high school students, and between 50% (Agnew et al., 1993) and 58% (Jones et al., 1992) of suitable behaviors being utilized by undergraduate students. Even at the graduate school level, Onwuegbuzie, Slate, and Schwartz (2001) found that graduate students enrolled in several sections of an introductory-level educational research course performed only 41% of the desirable study behaviors measured.

Although more than 50% of 4-year state institutions have implemented study skill programs (Cowart, 1987), mixed findings have been reported with respect to their ability to improve academic achievement (Kirschenbaum & Perri, 1982). As noted by Jones, Slate, and Marini (1995), before the potential for success of study skills programs can be maximized, knowledge of factors related to study skills must be improved. Indeed, Jones, Slate, and their colleagues have undertaken noteworthy work in this area. However, their subjects have tended to involve students from geographically-restricted areas. Thus, it

---

William J. Lammers is an associate professor in the Department of Psychology and Counseling at the University of Central Arkansas. Anthony J. Onwuegbuzie is an associate professor in the Department of Human Development and Psychoeducational Studies at Howard University. John R. Slate is a professor in the Department of Psychology and Counseling at the University of Texas at El Paso. Please address correspondence relating to the article to William J. Lammers, Department of Psychology and Counseling, UCA Box 4915, University of Central Arkansas, Conway, AR 72035-0001 or by e-mail to billl@mail.uca.edu.

cannot be assumed that their findings generalize to other regions.

In one study, Jones and his colleagues examined study skills as a function of gender, age, college classification (freshman, etc.), and grade point average (Jones et al., 1994). According to these researchers, female college students reported better study habits than did their male counterparts, classification was not related to study habits, and study habits were positively related to grade point average (i.e., GPA). However, the investigators did not examine the employment characteristics of the students. Surprisingly, study skills did not improve across classifications. Yet, one would have expected that students would learn better studying techniques as they progress from freshman to senior year. One possibility for this lack of relationship between study habits and year of study is that students also are spending more time in paid employment across their college years and that this negatively affects their study habits. Interestingly, in another study, Jones, Slate, and Marini (1995) did find a small ($r = -.15$) but statistically significant relationship between hours employed and study time but did not find any association between hours employed and study habits. However, participants in the latter study were from an introductory-level psychology course that was dominated by freshmen, with relatively few juniors and seniors. No other study was found in which the relationship between study habits and employment status was examined.

Employment status has become an increasingly relevant factor with the rising cost of a college education (U.S. Department of Education, 1997). Unfortunately for students, the cost of tuition has outpaced inflation over the last decade (Horn, 1998). Further, the total amount of scholarship money has decreased, whereas the use of loans has increased. Finally, increasing numbers of people are going to college who are also supporting families. As a result of these trends, more and more undergraduate students are engaged in paid employment (Horn, 1998). Indeed, at present, the majority of undergraduate students work while enrolled in college--with a significant proportion working on a full-time basis (Cuccaro-Alamin & Choy, 1998; Horn, 1994). Moreover, this paid employment is not limited to only part-time students, but to many full-time students as well (Horn, 1998).

Unfortunately, an analysis of the National Postsecondary Student Aid Study in 1996 revealed that working while enrolled in college tended to place limits on students' academic programs of study (Horn, 1998). For more than one-third of working students, these limitations involved restricting their choice of classes, reducing the number of classes that they could take at any particular time, limiting the time of day in which they could take classes, and minimizing their access to academic libraries. In fact, the more hours that students worked, the more likely they were to cite one of these four limitations. Specifically, whereas less than one-quarter of students working 15 or fewer hours per week reported that work imposed these limitations, more than 40% of students working full-time mentioned at least one of these limitations.

Additionally, Horn (1998) found that working more hours while enrolled in colleges was associated with higher rates of enrollment interruption. Specifically, students who worked between 1 and 15 hours per week, as well as those students who did not work at all, were less likely to interrupt their enrollment (i.e., have higher levels of persistence) than did those students who worked between 16 and 34 hours per week. Similarly, students who did not work at all tended to have the highest rates of persistence. This negative relationship between the number of hours worked and academic persistence still existed after adjusting for factors such as income, attendance status, and institution type. Indeed, approximately 20% of first-year students working full-time did not attend their colleges for a full year, compared with 5% of those students working 15 hours or less.

Even more importantly, according to Horn (1998), more than one-quarter of undergraduates who were employed reported that work adversely affected their levels of academic achievement. Indeed, a high negative relationship appears to exist between the number of hours that students work and their perceived levels of academic performance--with students at the high end of the work continuum (i.e., up to 34 hours) being more likely to report that working had a negative impact on their academic performance. Disturbingly, students who worked for at least 15 hours per week were more than twice as likely to indicate that their work had an adverse effect on their education.

Thus, little doubt exists that working while enrolled in college is a threat to students' levels of academic achievement. That is, college students who work long hours may be at an educational disadvantage compared to their non-working and low-working counterparts. Yet, little information is known about *how* working affects academic performance. However, it is likely that the relationship between the number of hours worked and academic achievement is moderated by students' study habits.

Thus, our purpose in conducting the present study was to examine simultaneously several variables that may affect academic achievement, namely, study habits, age, gender, classification, study time, and employment status. In many ways, we attempted to replicate and extend the findings of Jones, Slate, and their colleagues. The following seven specific research questions were addressed:

a) What are the academic strengths and weaknesses that are most prevalent? b) Are the study skills of students related to their overall academic achievement? c) How are college students' study skills related to their age? d) Are gender differences present in study skills? e) How are college students' study skills related to their classification? f) How are college students' study skills related to their time spent studying? and, g) How are college students' study skills related to their employment status?

Method

Participants

Research participants were 366 undergraduate students enrolled in nine different courses within the College of Education at a university in the midsouthern United States. Approximately 95% of the students were registered as full-time students, taking four or more courses per semester. Courses surveyed included General Psychology, Psychological Statistics, Experimental Psychology, Physiological Psychology, Child/Adolescent Development, Exceptionally and Culturally Diverse Student Population, History and Philosophy of Education, Applied Learning, and Classroom Behavior and Management. Participants included 82 males and 246 females (37 unknown) who represented 29 different majors on campus. The ages of the participants ranged from 17 to 49 years ($M = 21.7$, $SD = 4.6$). The estimated ethnic distribution, based on the population of students, was 87% White American, 12% African American, and 1% International.

Instruments and Procedure

The researchers selected a diverse set of courses within the College of Education. Students in these courses completed a questionnaire in which demographic information, including time spent studying and time spent working in paid employment, was elicited. Additionally, the questionnaire contained the Study Habits Inventory (SHI; Jones & Slate, 1992) that consists of 63 true-false items designed to assess typical study behaviors of college students. Thirty items describe effective study behaviors, and 33 items are related to ineffective study behaviors. The latter items were key-reversed such that total scale scores ranged from 0 to 63, with high scores indicating good study skills. This instrument has been found to generate reliable scores. For example, Jones and Slate (1992) reported SHI scores that had a classical theory mean alpha reliability coefficient of .85 and a 2-week test-retest coefficient of .82. The validity of scores on the SHI has been demonstrated through statistically significant correlations with college students' grades at both the undergraduate (Jones & Slate, 1992) and graduate (Onwuegbuzie et al., 2000) levels. As recommended by many researchers (e.g., Onwuegbuzie, 1999; Onwuegbuzie & Daniel, in press-a, in press-b; Thompson & Vacha-Haase, 2000; Wilkinson & the APA Task Force on Statistical Inference, 1999), reliability coefficients always should be reported for the data at hand. For the present study, scores pertaining to the SHI had a classical theory alpha reliability coefficient of .86.

Results

The means and standard deviations of all variables are presented in Table 1. In particular, the mean SHI score for students in this study was 33.36, indicating that they typically performed only 53.0% of appropriate behaviors that were assessed by the SHI. This mean is comparable to the means found in previous research of college students of 32.0 (Agnew et al., 1993), 33.0 (Jones, 1989), 33.7 (Jones, Slate, et al., 1993), 34.2 (Jones, Slate, et al., 1993; Jones et al., 1994), and 36.4 (Jones, Slate, & Kyle, 1992). The 95% confidence interval pertaining to the percentage of suitable behaviors in the present study was 51.5% to 54.5%. This relatively narrow interval suggests that students were homogeneous with respect to their study habits.

Table 1
Descriptive Statistics of Variables

Variable	M	SD	95% Confidence Interval Lower Limit	95% Confidence Interval Upper Limit
Grade Point Average	3.13	0.50	3.07	3.18
Hours/week studying	11.60	8.60	10.71	12.50
Hours/week working	15.60	14.20	14.11	17.13
Study Habits Score	33.36	9.25	32.41	34.31

As recommended by Jones, Slate, and colleagues (1992), study skill strengths were defined as those SHI items on which at least 75% of the students responded in an appropriate manner (i.e., responding "true" to items that described appropriate behaviors and "false" to items that characterized inappropriate behaviors). Conversely, study skill weaknesses were defined as those SHI items on which at most 25% of the students responded in an appropriate manner. This method resulted in the classification of 13 characteristic strengths within the sample and 7 characteristic weaknesses (see Table 2 and Table 3).

Table 2
Characteristic Strengths in Students' Study Skills

Study Habits Item	Percent Responding Appropriately
Note-Taking	
If I am sure I will remember something, I do not write it in my notes even if it seems to be important.	78.4
I take notes on odd, loose slips of paper instead of in a notebook.	91.0
I tape record lectures instead of taking notes.	98.1
I take notes after I have completed a reading assignment rather than taking notes as I go along.	80.9
Time Management	
I often do not have reports ready on time, or they are done poorly if I am forced to have them in on time.	86.3
I often sit down to study only to find that I do not have the necessary books, notes, or other materials.	82.0
If I plan to study with friends, I do not study by myself ahead of time.	77.6
Study Techniques	
I often try to make school more enjoyable by having a beer while I study.	92.9
In studying a textbook, I try to memorize the exact words in the text.	79.2
I try to break large amounts of information into small clusters that can be studied separately.	81.1
I use the facts I learned in one course to help me understand the material in another course.	85.0
I use the facts learned in school to help me understand events outside of school.	84.4
I try to think critically about new material and not simply accept everything read	77.9

Table 3
Characteristic Weakness in Students' Study Skills

Study Habits Item	Percent Responding Appropriately
Note-Taking	
As soon as possible after class, I recopy my lecture notes.	9.0
I keep a special indexed notebook or card system or recording new words and their meanings.	8.5
Reading Skills	
I use the headings to make an outline of a chapter before I begin to read it.	22.1
Before reading a chapter, I jot down a few questions and a list of key terms to focus my attention while reading.	9.8
Sometimes I discover that I have "read" several pages without knowing what was on them.	11.7
Sometimes I make simple charts or diagrams to show how the facts I am learning are related to each other.	20.5
Time Management	
I do most of my reviewing for a test the night before the examination.	19.7

Study Skills Strengths. A content analysis of the identified strengths indicated that these strengths fell into the following three categories: note-taking, time management, and study techniques (Table 2). With respect to taking notes, students tended to report that they used notebooks rather than loose paper to take notes. These students also were unlikely to use a tape recorder as a replacement for, rather than as an adjunct to, taking notes. Additionally, they tended to take notes as they read material, rather than waiting until they had completed their reading assignments to take notes. Finally, these students took notes on any material that they deemed to be important, even if they were confident that they would remember it.

With respect to time management, students were likely to have papers completed on time. They typically had the necessary materials to study, and when they planned to study with their peers, they tended to study by themselves ahead of time. When studying, students with appropriate study techniques tended to avoid consumption of alcoholic beverages. They also tried not to rely on rote memorization, preferring to relate course materials to everyday life and to material in other courses. These students also were more apt to break down the study material into meaningful components that could be studied separately, and to reflect on new material rather than accepting everything to which they were exposed.

Study Skills Weaknesses. A content analysis of weaknesses revealed general themes related to a) note-taking, b) reading skills, and c) time management. With respect to the former, students tended not to use designated notebooks to record new words and their meanings, nor did they recopy their lecture notes. Interestingly, reading was the students' weakest area of academic skill. Indeed, four weaknesses were identified. Specifically, students reported that they did not preview chapters of a book before reading them by creating outlines. Neither did students develop simple charts or diagrams that illustrated how material was inter-connected, nor did students make lists of key terms to help them focus while reading. Consistent with their passive reading styles was the fact that the majority of students reported that they often "read" several pages without knowing what was on them. Unfortunately, students reported a major time management deficiency, namely, waiting until the night before an examination to undertake the bulk of their studying.

Intercorrelations of Variables

The Pearson Product-Moment correlation coefficients among age, GPA, hours per week spent studying, hours per week spent working, and the SHI scores are presented in Table 4. After applying the Bonferroni adjustment, a statistically significant relationship was found between age and the number of hours spent studying per week. Specifically, older students tended to spend more time studying. Using Cohen's (1988) criteria, the effect size associated with this relationship was small to moderate. Interestingly, a significant and moderate negative relationship was found between students' GPA and the number of hours that they spent working each week, indicating that students who spent the most time in paid employment tended to have the lowest levels of academic achievement. Of most interest was the fact that the SHI scores were related positively to age, GPA, and the number of hours spent studying each week, and related negatively to the number of hours spent working each week. The relationship with the largest effect size was that between SHI scores and GPA.

Table 4
Pearson Product-Moment Correlations Among, Age, GPA, Hours of Study Per Week, Hours of Work Per Week, and Study Habits Scores

	Age	GPA	Hours/week studying	Hours/week working
GPA	.10			
Hours/week studying	.25*	.12		
Hours/week Working	.06	-.28*	-.15*	
Study Habits Score	.19*	.36*	.24*	-.16*

* Statistically significant after applying the Bonferroni adjustment

A series of t-tests also was conducted comparing males and females with respect to age, GPA, hours per week spent studying, hours per week spent working, and the SHI scores. After using the Bonferroni adjustment, no statistically significant gender difference was found with respect to age ($t = -.06$, $df = 187.90$, $p > .05$) and the number of hours spent working per week ($t = 1.93$, $df = 121.01$, $p > .05$). However, (a) females ($M = 11.79$, $SD = 8.83$) were found to spend significantly ($t = 2.94$, $df = 199.73$, $p < .05$; Cohen's (1988) $d = 0.31$) more time studying per week than did males ($M = 9.20$, $SD = 6.01$); (b) females ($M = 3.17$, $SD = 0.49$) were found to report significantly ($t = 2.46$, $df = 135.32$, $p < .05$; Cohen's $d = 0.32$) higher GPAs than did males ($M = 3.01$, $SD = 0.50$); and (c) females ($M = 34.46$, $SD = 9.68$) were found to report significantly ($t = 5.13$, $df = 183.12$, $p < .05$; Cohen's $d = 0.57$) higher SHI scores than did their male counterparts ($M = 29.26$, $SD = 7.29$). Using Cohen's (1988) criteria, the effect sizes pertaining to these differences were moderate to large.

An analysis of variance (ANOVA) was undertaken to examine age, GPA, hours per week spent studying, hours per week spent working, and the SHI scores as a function of college classification level. Findings revealed no difference in GPA across classifications [$F(3, 341) = 2.09$, $p > .05$, $\omega^2 = 0.14$], no difference in hours spent studying per week across classifications [$F(3, 346) = 0.38$, $p > .05$, $\omega^2 = 0.05$], and no difference in SHI scores across classifications [$F(3, 353) = 1.41$, $p > .05$, $\omega^2 = 0.11$]. However, the hours spent working per week discriminated the four groups [$F(3, 330) = 7.30$, $p < .05$, $\omega^2 = 0.26$]. The effect size associated with this difference was moderate. A Scheffé post-hoc analysis revealed that Juniors ($M = 16.42$, $SD = 13.79$) and Seniors ($M = 18.36$, $SD = 14.04$) spent more hours working per week than did Freshmen ($M = 9.25$, $SD = 11.24$). The hours spent working per week by Sophomores ($M = 13.81$, $SD = 13.18$) was not significantly different than that for the other classifications. Moreover, an examination of polynomial contrasts revealed a linear trend [$F(1, 353) = 21.15$, $p < .05$], with the number of hours worked by week increasing monotonically as a function of year of study. Consistent with this finding was the fact that the percentage of students employed increased as classification increased, $\chi^2(3) = 11.53$, $p < .05$. Cramer's V associated with this relationship was .18, which suggests a small-to-moderate effect size.

Regression Analysis Using GPA as the Dependent Variable

The ability of several variables to predict academic success was assessed using an *All possible subsets* (APS) multiple regression (Tabachnick & Fidell, 1996; Thompson, 1995). College GPA was the criterion variable, and predictor variables were SHI score, hours spent working per week, hours spent studying per week, age, gender, and classification (i.e., freshman vs. non-freshman). APS multiple regression was utilized to select an optimal set of variables in terms of maximum proportion of variance explained by the predictor set. All possible models involving some or all of the selected variables were examined (Tabachnick & Fidell, 1996). This method of analysis has been recommended by many statisticians (e.g., Thompson, 1995). Indeed, APS multiple regression has been found to be superior to stepwise multiple regression in finding the optimal model (Huberty, 1989; Thompson, 1995; Thompson, Smith, Miller, & Thomson, 1991). In APS regression, separate regressions are computed for all independent variables singly, all possible pairs of independent variables, all possible trios of independent variables, and so forth, until the best subset of independent variables is identified according to some criterion. For the present study, the criterion used for selection of the optimal APS multiple regression model was the maximum proportion of variance explained (R^2), which provides an important measure of effect size (Cohen, 1988). An examination of skewness and kurtosis coefficients did not indicate that the distribution of GPA scores was non-normal, thereby justifying the use of multiple regression.

In addition, evaluation of assumptions of linearity and homogeneity revealed no threat to multiple regression analysis.

The analysis indicated that SHI score ($\beta = .33$) and hours spent working per week ($\beta = -.22$) were significant predictors of GPA, $F(2, 286) = 31.32, p < .001$. These two variables combined to explain 18.0% of the total variance (adjusted $R^2 = 17.4\%$). The SHI score explained 13.2% of the variance in GPA, and hours working explained an additional 4.8% of the variance. (Analysis of residuals suggested no departure from normality). Using Cohen's (1988) criteria for multiple regression models, the effect size pertaining to SHI was moderate, whereas that pertaining to hours working was small. The model indicates that the higher the GPA, the better the study skills students displayed and the fewer hours they spent working per week.

Because SHI score was a significant predictor of GPA, an additional regression analysis was undertaken, using GPA as the criterion variable and the SHI items as the predictor variables. Such an analysis was performed by Jones et al. (1995). Despite the fact that all SHI items were examined, the subject-to-variable ratio still exceeded the 5 to 1 minimum recommended for multivariate analyses in general (Thompson, 1990) and multiple regression in particular (Tabachnick & Fidell, 1996). The selected multiple regression model contained eight variables that significantly predicted GPA [$F(8, 328) = 12.49, p < .001$]. These variables combined to explain 23.4% of the total variance (adjusted $R^2 = 21.5\%$). The eight SHI items that contributed to the regression equation are presented in Table 5. It can be seen from this table that students with the highest GPA were more likely than those students with the lowest GPA to concentrate while studying for short periods, to study beyond the point of immediate recall, to recopy their lecture notes shortly after class, to complete reports on time, to identify important aspects of the material they read, to seek help from their instructors when needed, and to record all important information in their notebooks. On the other hand, students with the highest GPAs were less likely than were their lower-achieving counterparts to have a special system for recording new words and their meanings.

Regression Analysis Using Total SHI Scores as the Dependent Variable

Because the total SHI score was the best predictor of GPA, an APS multiple regression analysis also was undertaken using total SHI score as the dependent variable, and GPA, hours spent working per week, hours spent studying per week, age, gender, and classification (i.e., freshman vs. non-freshman) as the criterion variables, as undertaken by Jones, Slate, and Marini (1995). Again, all model assumptions appeared to hold. The selected regression model revealed that GPA ($\beta = .33$), the number of hours spent studying per week ($\beta = .17$), and gender ($\beta = .14$) significantly predicted SHI scores [$F(3, 285) = 21.50, p < .001$]. These variables combined to explain 18.5% of the total variance (adjusted $R^2 = 17.6\%$). Specifically, GPA explained 13.2% of the variance, with the number of hours spent studying per week and gender explaining 3.5% and 1.8% of the variance, respectively. The model indicates that study skill use tended to improve with GPA, the number of hours spent studying per week, and female status.

Table 5
Study Habits that Contributed Significantly to the Prediction of GPA

Study Habits Item	Beta	% of Variance Explained
My study periods are too short for me to get "warmed up" and really concentrate on studying.	.22	8.0
I try to do some "overlearning," working beyond the point of immediate recall.	.17	4.7
As soon as possible after class, I recopy my lecture notes.	.16	2.9
I often do not have reports ready on time, or they are done poorly if I am forced to have them in on time.	.11	2.5
I have trouble in picking out the important points in the material I read.	.13	1.8
I keep a special indexed notebook or card system for recording new words and their meanings.	-.13	1.4
When I have difficulty with my work, I do not hesitate to seek help from my instructor.	.11	1.0
If I am sure I will remember something, I do not write it in my notes even if it seems to be important.	.11	1.1

Note: Positive βs indicate that students with the highest GPA tended to report more appropriate behavior; negative βs indicate that students with the highest GPA tended to report more negative behaviors.

Discriminant Analyses

GPA. As undertaken on several occasions by Jones, Slate, and their colleagues (e.g., Jones, Slate, & Kyle, 1992; Jones et al., 1994; Jones, Slate, & Marini, 1995), students in the upper third of the GPA distribution ($n = 120$, range = 3.40-4.00) were contrasted with students in the lower third of the GPA distribution ($n = 104$, range = 1.60-2.90). A canonical discriminant analysis was then undertaken comparing the low and high GPA groups, using the individual SHI items as the discriminating variables (Tabachnick & Fidell, 1996). Although this analysis resulted in data loss (i.e., the middle GPA group removed), contrasting the upper and lower group often results in adequate statistical power because of the removal of misclassifications near the median that are due to measurement error (Dane, 1990; Jones, Slate, & Marini, 1995).

The resulting discriminant function was statistically significant, χ^2 (63) = 122.39, $p < .001$, and accounted for 49.0% of the between groups variance (canonical R = .70). The group centroids were .92 for high achievers and -1.04 for low-achieving students, indicating that this function primarily discriminated low- and high-achieving students. An examination of the pooled-within-group correlations indicated that, using a cutoff loading of 0.3 (Lambert & Durand, 1975; Tabachnick & Fidell, 1996), five items made an important contribution to the canonical function. These items are listed in Table 6. The positive correlation coefficients indicate that high-achievers were more likely to respond appropriately to this item (i.e., more likely to report an appropriate behavior), whereas negative correlation coefficients indicate that low-achieving students were more likely to respond appropriately to this item. Thus, Table 6 indicates that high achievers were more likely to concentrate while studying for short periods, to identify the important points in the material read, and to study by themselves ahead of time when they planned to study with their peers. Also, high achievers were less likely to daydream when sitting in classes and to loaf when they should be studying.

Table 6
Study Habits Inventory Items as a Function of GPA and Gender, with Pooled-Within-Subjects Correlations of .30 or Greater in the Discriminant Analysis

Variable	Study Habits Item	r
GPA		
	My study periods are too short for me to get "warmed up" and really concentrate on studying.	.38
	I spend too much time on loafing, movies, dates, and so forth that I should be spending on my course work.	.31
	I have trouble in picking out the important points in the material I read.	.31
	If I plan to study with friends, I do not study by myself ahead of time.	.30
	When sitting in my classes, I have a tendency to daydream about other things.	.30
Gender		
	If I am sure I will remember something, I do not write it in my notes even if it seems to be important.	.44
	I spend too much time on loafing, movies, dates, and so forth that I should be spending on my coursework.	.32
	I usually write reports several days before they are due, so that I can correct them if necessary.	.33

Age. A canonical discriminant analysis was used to contrast students in the upper third of the age distribution (n = 128, range = 22-49 years) with students in the lower third of this distribution (n = 105, range = 17-20 years). The discriminant function was statistically significant, χ^2 (63) = 106.69, $p < .001$, and accounted for 42.3% of the between groups variance (canonical R = .65). The group centroids were .79 for older students and -0.94 for younger students. Interestingly, however, no SHI items had pooled-within-groups correlations above .30.

Gender. A canonical discriminant analysis was used to discriminate males (n = 82) from females (n = 246). The discriminant function was statistically significant, P^2 (63) = 140.08, $p < .001$, and accounted for 39.7% of the between groups variance (canonical R = .63). The group centroids were -1.41 for males and .46 for females. Three SHI items made an important contribution to the canonical function. These items are presented in Table 6. The positive correlation coefficients indicate that females were more likely to respond appropriately to this item, whereas negative correlation coefficients indicate that males were more likely to respond appropriately to this item. It can be seen from Table 6 that female students were more likely to take notes of any material that was deemed important and to complete assignments several days before they were due to provide enough time to make necessary revisions, and were less likely to loaf when they should have been studying.

Hours Spent Studying. Similarly, a canonical discriminant analysis was used to contrast students in the upper third with respect to the number of hours spent studying per week (n = 133, range = 12-80 hours) with students in the lower third with respect to study time (n = 133, range = 0-7 hours). The discriminant function was statistically significant, χ^2 (63) = 98.01, $p < .005$, and accounted for 38.4% of the between groups variance (canonical R = .62). The group centroids were 0.74 for those students who spent the most amount of time studying and -0.85 for students who spent the least amount of time studying. The six items that made an important contribution to the canonical function are presented in Table 7. The positive correlation coefficients indicate that those students who studied the most were more likely to respond appropriately to this item. Table 7 reveals that students who spent the most amount of time studying were more likely than were those students who spent the least amount of time studying per week to study with the intent of retaining material on a long-term basis, not to skip classes in which attendance is optional, to have specific schedules for each subject, to review well before examinations are due, to exhibit much less difficulty getting down to work and maintaining attention to their work.

Table 7
Study Habits Inventory Items as a Function of Hours Spent Studying Per Week and Hours Spent Working Per Week, with Pooled-Within-Subjects Correlations of .30 or Greater in the Discriminant Analysis

Variable	Study Habits Item	r
Hours Spent Studying		
	I study most subjects with the idea of remembering the material only until the test is over.	.31
	I sometimes skip classes, especially when attendance is not required.	.30
	I have a definite, although reasonably flexible, study schedule with times for studying specific subjects.	.33
	I spend too much time on loafing, movies, dates, and so forth that I should be spending on my course work.	.39
	I do most of my reviewing for a test the night before the examination.	.37
	I have to wait for the mood to strike me before attempting to study.	.40
Hours Spent Working		
	I sometimes skip classes, especially when attendance is not required.	.31
	My study periods are too short for me to get "warmed up" and really concentrate on studying.	.33
	I frequently do not get enough sleep and feel sluggish in class or when studying.	.31
	I review frequently.	.33

Hours Spent Working. Finally, a canonical discriminant analysis was used to contrast students in the upper third with respect to the number of hours spent working per week ($n = 116$, range = 23-60 hours) with students in the lower third with respect to work hours ($n = 115$, no hours spent working). The discriminant function was statistically significant, χ^2 (63) = 85.99, $p < .05$, and accounted for 36.0% of the between groups variance (canonical $R = .60$). The group centroids were -0.75 for those students who spent the most amount of time working and 0.75 for students who spent the least amount of time working. The four items that made an important contribution to the canonical function are presented in Table 7. The positive correlation coefficients indicate that those students who spent the most time working per week were less likely to respond appropriately to this item. Table 7 reveals that students who spent the most amount of time working were less likely than their counterparts to review material frequently, to concentrate on studying, to not be sleep deprived, and to attend classes in which attendance was optional.

Discussion

As is the case in previous studies of undergraduate students (e.g., Agnew et al., 1993; Jones, 1989; Jones et al., 1993; Jones et al., 1994; Jones, Slate, & Kyle, 1992; Jones, Slate, & Marini, 1995), the present sample exhibited poor study skills, performing appropriately only slightly more than one-half (i.e., 53%) of the study behaviors assessed on the SHI. These findings, together with the moderate positive relationship between SHI scores and academic achievement found in the present study and in previous investigations (Agnew, Slate, Jones, & Agnew, 1993; Elliot et al., 1990; Jones, Slate, & Kyle, 1992; Jones, Green et al., 1993; Jones et al., 1994; Jones, Slate, & Marini, 1995; Kleijn, van der Ploeg & Topman, 1994), suggest the importance of study skills training programs. However, it is likely that study skills training by itself will not be effective unless the instruction is combined with effective motivational techniques (Brophy, 1987; Jones, Slate, & Kyle, 1992). Moreover, according to some researchers (e.g., Langer & Neal, 1987), study skills training appears to be most effective when domain-specific skills are taught as part of the courses that students are currently taking.

As noted by Jones, Green et al. (1993), it is inefficient for study skills programs to devote too much time, money, and resources to teaching skills that are already practiced by the majority of students. Rather these programs should focus more on typical study skills weaknesses. As such, the present study has contributed to the knowledge base by identifying some of these deficiencies.

The most frequently cited weaknesses involved note-taking, time management, and reading skills--with behaviors comprising the latter being the most predominant. The fact that the majority of students read material in a passive manner is a serious cause for concern because it reduces students' ability to understand and to interpret lecture notes, books, and the like. In fact, it should be no surprise that nearly 90% of students reported that they can read several pages of a book without understanding its content. Interestingly, low-achieving college students tend to spend more of their study time simply reading and rereading study materials than do high achievers (Rasheed, O'Neill, Walters, & Johnston, 1975). Even if the material that is read is understandable to the student, passive readers often find it difficult to retain what they read for a significant length of time (Jones, Slate, & Kyle, 1992).

The result that less than 10% of students made any special effort to learn new terminology (e.g., by recording new words and their meanings) may help to explain further the relationship between study skills and academic achievement. Interestingly, Rasheed et al. (1975) found that unsuccessful students tend to spend less time learning new terms than do their more successful counterparts.

Another major study skill weakness was waiting until the last moment to study for an examination. This weakness has been identified in other studies (e.g., Jones et al., 1991; Slate et al., 1993; Stanley, Slate, & Jones, 1999). As noted by Stanley et al. (1999), this behavior is a particularly important weakness because the ability to eradicate this weakness depends, in part, on how much study time is available. In fact, an examination of point biserial correlations (not presented) revealed that this weakness was more strongly related to the number of hours spent studying per week ($r = .25, p < .001$) than to any other study skill deficit.

Findings from the present study also suggest that students' use of appropriate study skills varies as a function of a number of factors, which include age, gender, number of hours spent studying each week, and the number of hours spent working each week. As such, these and other factors identified by Jones, Slate, and their colleagues should be addressed as part of the design of study skills training programs.

The fact that older students tended to exhibit better overall study skills than did their younger counterparts is consistent with Jones et al. (1994) and Jones, Slate, and Marini (1995). This finding suggests that study skills are a function of maturity level. Ironically, many older adults think that they lack the academic skills necessary to maintain the same level of academic achievement as younger students (Schlossberg, Lynch, & Chickering, 1989). Unfortunately, the discriminant function, although statistically significant, did not lead to the identification of specific study behaviors that discriminated younger and older students. Thus, the role of age in the acquisition of study skills warrants further investigation.

Although no gender differences were found with respect to the number of hours spent working per week, females reported spending significantly more time studying. This finding suggests that females devoted more of their non-work time for academic purposes. This gender difference in the amount of time spent studying may explain, at least in part, why females had higher SHI scores, as well as higher levels of academic achievement. A discriminant analysis revealed that females were more likely than were males to take notes of any material that was deemed important and to complete assignments several days before they were due to provide enough time to make necessary revisions and were less likely to loaf when they should have been studying. Thus males may benefit from learning time management and note-taking strategies.

The fact that study skills were positively related to time spent studying is in accordance with Jones, Slate, and Marini (1995). This relationship has intuitive and logical appeal because it is likely that students who spend the least time studying have little time to utilize effective study habits. Indeed, a discriminant analysis revealed that students who spent the least amount of time studying per week were more likely to have inconsistent study patterns and to undertake most of their reviews for examination the night before they took place. Furthermore, it is clear that hours spent studying is based on motivational factors because students who spent the least amount of time studying tended to report that they spent more time loafing and engaging in non-academic activities than did their counterparts, and that they had to "wait for the mood to strike" before attempting to study. This latter finding lends support to the contention (e.g., Jones, Slate, & Marini, 1995) that study skills programs also should address motivational issues.

Study skills were related to the number of hours spent working per week. Specifically, those students who spent the most time working per week were less likely than were their counterparts to attend class regularly, to review material frequently, to exhibit much less difficulty getting down to work, and not to be sleep deprived. These students might benefit from motivational and time management strategies. However, most importantly, they should be made aware that working while in college has been related to lower academic performance (Horn, 1998).

That study habits scores did not change as a function of classification is somewhat a cause for concern because one would expect that students would develop better study habits as they progress through college. Similarly, one would have expected the number of hours spent studying per week and GPA to increase with classification level. However, it appears that the number of hours worked attenuates these relationships, thereby acting as a mediating variable. That is, although seniors may possess more knowledge of appropriate study skills and may have more academic motivation than do freshmen, working more hours per week prevents them from spending more time studying and, consequently, from exhibiting better study skills. This, in turn, diminishes their levels of academic achievement, thereby creating an equalizing effect.

Nevertheless, it is likely that many of the variables examined in the present study (e.g., study skills, the number of hours worked per week, and academic achievement) are intricately intertwined. For example, even if students have knowledge of the appropriate study skills, their ability to utilize this knowledge will probably be affected by the number of hours per week that they work. The fewer hours that a student works, the more time that a student has to put these appropriate study skills into operation. Additionally, for students with poor study skills, working many hours while at college may not afford them the time to develop appropriate skills, which includes reducing the number of hours spent studying per week--as evidenced by the relationship found between the hours spent working and the hours spent studying. Educators and academic advisors need to be aware of how the number of hours worked per week by students may affect the success of study skills programs.

An important limitation of the present study is that data were collected from college students in a geographically-restricted region. Thus, it is not clear the extent to which these findings generalize to students from other geographic regions, suggesting a need for replication using diverse samples and examining other variables. Additionally, the current findings stem from a correlational research design. Thus, researchers in the future should examine the causal nature of some of the relationships found in this study.

References

Agnew, N. C., Slate, J. R., Jones, C. H., & Agnew, D. M. (1993). Academic behaviors as a function of academic achievement, locus of control, and motivational orientation. *NACTA Journal, 37*, 24-27.

Brophy, J. (1987). Synthesis of research on strategies for motivating students to learn. *Educational Leadership, 45*, 40-48.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: John Wiley.

Cowart, S. C. (1987). *What works in retention at state colleges and universities.* Iowa City, IA: The ACT National Center for the Advancement of Educational Practices, The American College Testing Program.

Cuccaro-Alamin, S., & Choy, S. (1998). *Post secondary financing strategies: How undergraduate combine work, borrowing, and attendance (NCES 98-088).* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Dane, F. C. (1990). *Research methods.* Pacific Grove, CA: Brooks/Cole.

Elliot, T. R., Godshall, F., Shrout, J. R., & Witty, T. E. (1990). Problem-solving appraisal, self-reported study habits, and performance of academically at-risk college students. *Journal of Counseling Psychology, 37*, 203-207.

Horn, L. J. (1994). *Undergraduates who work while enrolled in postsecondary education (NCES 94-311).* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Horn, L. J. (1998). *Undergraduates who work: National postsecondary student aid study (NCES 98-137).* Washington, DC: U.S. Government Printing Office.

Huberty, C.J. (1989). Problems with stepwise methods--better alternatives. In B. Thompson, (Ed.), *Advances in social science methodology,* Vol 1. (pp. 43-70). Greenwich, CT: JAI.

Jones, C.H. (1989, Spring). Improving students' study skills. *Arkansas School Psychology Association Newsletter*, pp. 4-5.

Jones, C. H., Green, A. E., Mahan, K. D., & Slate, J. R. (1993). College students' learning styles, academic achievement, and study behaviors. *Louisiana Education Research Journal, 19*, 40-48.

Jones, C.H., & Slate, J.R. (1992). *Technical manual for the Study Habits Inventory.* Unpublished manuscript, Arkansas State University, Jonesboro, AR.

Jones, C. H., Slate, J. R., & Kyle, A. (1992). Study skills of teacher education students. *The Teacher Educator, 28*, 7-15.

Jones, C. H., Slate, J. R., Bell, S., & Saddler, C. (1991). Helping high school students improve their academic skills: A necessary role for students. *The High School Journal, 74*, 198-202.

Jones, C. H., Slate, J. R., Blake, P., & Holifield, S. (1992). Two investigations of the academic skills of junior and senior high school students. *The High School Journal, 76*, 24-29.

Jones, C. H., Slate, J. R., Mahan, K. D., Green, A. E., Marini, I., & DeWater, B. K. (1994). Study skills of college students as a function of gender, age, class, and grade point average. *Louisiana Education Research Journal, 19*, 60-74.

Jones, C. H., Slate, J. R., & Marini, I. (1995). Locus of control, social interdependence, academic preparation, age, study time, and the study skills of college students. *Research in the Schools, 2*, 55-62.

Jones, C. H., Slate, J. R., Marini, I., & DeWater, B. K. (1993). Academic skills and attitudes toward intelligence. *Journal of College Student Development, 34*, 422-424.

Jones, C. H., Slate, J. R., Perez, E., & Marini, I. (1996). Graduate students' study skills as a function of academic achievement, sex, conceptions of intelligence, and locus of control. *New Directions for Education Reform, 3*, 61-78.

Kirschenbaum, D. S., & Perri, M.G. (1982). Improving academic competence in adults: A review of the recent research. *Journal of Counseling Psychology, 25,* 76-94.

Kleijn, W. C., van der Ploeg, H. M., & Topman, R. M. (1994). Cognition, study habits, test anxiety, and academic performance. *Psychological Reports, 75,* 1219-1226.

Lambert, Z., & Durand, D. (1975). Some precautions in using canonical analysis. *Journal of Market Research, XII,* 468-475.

Langer, M. A., & Neal, J. C. (1987). Strategies for learning: An adjunct study skills model. *Journal of Reading, 31*, 134-139.

Lawler-Prince, D., Slate, J. R., & Jones, C. H. (1993). Academic behaviors of preservice elementary and/or special education teachers: A preliminary study. *Louisiana Educational Research Journal, 18*, 109-118.

Onwuegbuzie, A. J. (in press). Common analytical and interpretational errors in educational research: An analysis of the 1998 volume of the British Journal of Educational Psychology. *Research for Educational Reform.*

Onwuegbuzie, A. J., & Daniel, L. G. (in press). Uses and misuses of the correlation coefficient. *RESEARCH IN THE SCHOOLS, 9*(1).

Onwuegbuzie, A. J., & Daniel, L. G. (in press). Typology of analytical and interpretational errors in quantitative and qualitative educational research. *Current Issues in Education.*

Onwuegbuzie, A. J., Slate, J. R., & Schwartz, R. A. (2001). Role of study skills in graduate-level educational research courses. *Journal of Educational Research, 94*(4), 238-246.

Onwuegbuzie, A. J., Slate, J. R., Paterson, F., Watson, M., & Schwartz, R.A. (2000). Factors associated with underachievement in educational research courses. *RESEARCH IN THE SCHOOLS,* 7(1), 53-65.

Rasheed, J.A., O'Neill, G.W., Walters, W.M., & Johnston, J.M. (1975). A description of typical study behaviors in a behaviorally taught course. In J.M. Johnston (Ed.), *Behavior research and technology in higher education* (pp. 421-435). Springfield, IL: Charles C. Thomas.

Schlossberg, N.K., Lynch, A.Q., & Chickering, A.W. (1989). *Improving higher education environments for adults: Responsive programs and services from entry to departure*. San Francisco: Jossey-Bass.

Slate, J. R., Jones, C. H., & Charlesworth, J. R. (1990). Relationship of conceptions of intelligence to preferred teaching behaviors. *Action in Teacher Education, 12*(1), 25-29.

Slate, J. R., Jones, C. H., & Dawson, P. (1993). Academic skills of high school students as a function of grade, gender, and academic track. *The High School Journal, 76*, 245-251.

Stanley, B., Slate, J. R., & Jones, C. H. (1999). Study behaviors of college preparatory and honors students in the ninth grade. *The High School Journal, 82*, 165-171.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper & Row.

Thompson, B. (1990). Finding a correction for the sampling error in multivariate measures of relationships: A Monte Carlo study. *Educational and Psychological Measurement, 50,* 15-31.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement, 55,* 525-534.

Thompson, B., Smith, Q. W., Miller, L. M., & Thomson, W.A. (1991, January). *Stepwise methods lead to bad interpretations: Better alternatives*. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60*, 174-195.

U.S. Department of Education. (1997). *Digest of Education Statistics 1996 (NCES 96-133)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.

Copyright 2001 by the
Mid-South Educational Research Association

RESEARCH IN THE SCHOOLS
2001, Vol. 8, No. 2, 83-91

The Utility of Statistical Significance Testing in Psychological and Educational Research: A Review of Recent Literature and Proposed Alternatives

Jeremy R. Sullivan
Texas A&M University

The issues and criticisms surrounding the use and utility of statistical significance tests have culminated in the recent work by the APA Task Force on Statistical Inference. Members of this committee put forth a set of guidelines regarding the use of statistical tests; these guidelines will likely appear in the next edition of the APA's **Publication Manual**. *It is therefore crucial that psychologists and educators are familiar with and understand these issues, especially with regard to the limitations of statistical tests. The present paper summarizes the post-1994 literature in psychology and education regarding statistical significance testing, with an emphasis on (a) both limitations and defenses of statistical testing and (b) proposed alternatives or supplements to statistical significance testing. It is concluded that responsible use of statistical tests entails recognition of the tests' limitations, in addition to the supplemental use of alternative analytic techniques such as effect sizes, confidence intervals, and replicability analyses.*

Researchers have long placed a premium on the use of statistical significance testing, notwithstanding withering criticisms of many conventional practices as regards statistical inference (e.g., Burdenski, 1999; Carver, 1978; Daniel, 1998; McLean & Ernest, 1998; Meehl, 1978; Morrison & Henkel, 1970; Nix & Barnette, 1998; Thompson, 1993, 1998a, 1998b, 1998c, 1999a, 1999b, 1999c). A series of articles on these issues appeared in recent editions of the *American Psychologist* (e.g., Cohen, 1990; Kupfersmid, 1988; Rosnow & Rosenthal, 1989). Especially noteworthy and influential are recent articles by Cohen (1994), Kirk (1996), Schmidt (1996), and Thompson (1996). Nickerson (2000) provides a comprehensive review of this literature, while Anderson, Burnham, and Thompson (2000) and Hubbard and Ryan (2000) empirically document related trends. Kirk (2001), Shibley Hyde (2001), and Vacha-Haase (2001) describe in detail the issues surrounding forthcoming changes in the APA publication manual and journal editorial policies with regard to the use and interpretation of statistical significance tests.

Indeed, the criticism of statistical testing is growing fierce. For example, Tryon (1998) recently lamented in the *American Psychologist*,

> The fact that statistical experts and investigators publishing in the best journals cannot consistently interpret the results of these analyses is extremely disturbing. Seventy-two years of education have resulted in minuscule, if any, progress toward correcting this situation. It is difficult to estimate the handicap that widespread, incorrect, and intractable use of a primary data analytic method has on a scientific discipline, but the deleterious effects are doubtless substantial . . . (p. 796)

Schmidt and Hunter (1997), virulent critics of statistical significance testing, similarly argued that, "Statistical significance testing retards the growth of scientific knowledge; it *never* makes a positive contribution" (p. 37, emphasis added).

Criticisms of the statistical significance testing procedure are prevalent and occur across many scientific disciplines. To be sure, this debate is not an esoteric one for pure statisticians to resolve; applied psychological, educational, medical, and other social science researchers and even clinicians have taken sides and argued their points cogently (Krantz, 1999; Svyantek & Ekeberg, 1995; Zakzanis, 1998). Indeed, a recent empirical study of four disciplines on a decade-by-decade basis found an exponential increase in criticisms across disciplines of statistical testing practices (Anderson et al., 2000). These criticisms are not only ubiquitous, but also are far from new (see Boring, 1919).

Jeremy R. Sullivan is a doctoral student in the school psychology program within the Department of Educational Psychology at Texas A&M University. He is employed as the graduate research assistant to the Dean of Education at Texas A&M. An earlier version of this paper was presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX, January 29, 2000. The author expresses his gratitude to Bruce Thompson and to four anonymous reviewers, whose comments and suggestions led to a more thoughtful and comprehensive paper. Address correspondence regarding this article to Jeremy R. Sullivan, Texas A&M University, College of Education, College Station, TX 77843-4222; or by e-mail to jeremy-sullivan@neo.tamu.edu.

The older commentary eventually led to a very important change in the 1994 APA publication manual: an "encouragement" (p. 18) to always report effect sizes. Further, the recently published report of the APA Task Force on Statistical Inference states that effect sizes should *always* be reported for *all* primary results (Wilkinson & APA Task Force on Statistical Inference, 1999). In fact, editors of several journals in psychology and education now explicitly *require* the reporting of effect sizes along with statistical test results (e.g., Heldref Foundation, 1997; McLean & Kaufman, 2000; Murphy, 1997; Thompson, 1994a). Admonition by the APA and journal editors to report effect sizes reflects growing discontent with the current state of affairs regarding the use of statistical significance tests and is indicative of a field ready for change.

The present paper explores these views in detail with a review of the literature in the fields of psychology and education, in addition to a discussion of proposed alternatives or supplements to statistical significance testing. Specifically, the criticisms against and limitations of statistical significance testing included in the present discussion are: (a) these analyses tell us nothing about result replicability, (b) these tests are heavily influenced by sample size, (c) statistical testing typically does not provide researchers with the information that they really want, and (d) statistical tests do not allow researchers to interpret the practical significance of their results. Also considered in the present discussion are defenses of statistical significance testing, including many researchers' beliefs that: (a) statistical tests are especially useful in testing ordinal claims, (b) superior alternatives to statistical testing are not readily available, and (c) the ubiquitous misuse of statistical tests is not the fault of the tests, but of the researchers who misuse them.

Due to the long history of these issues, arguments, and criticisms of statistical testing, the present review will emphasize the most recent (i.e., post-1994) literature in this area. Recognizing that the issues presented here have been tackled elsewhere, the present paper will attempt to provide an integrated and unique contribution that is nontechnical and conceptually basic, in the hope that all readers, from novice statistician to seasoned statistical veteran, will find the coverage interesting, enlightening, and accessible.

Defenses of Statistical Significance Testing

A perusal of some of the most popular journals in education and psychology would likely indicate that statistical significance testing has seemingly withstood all the criticism, as it remains a widely-used analytical tool in these fields (Loftus & Masson, 1994; Shrout, 1997). This section addresses several reasons why statistical testing has weathered the storm, and why many researchers continue to use statistical significance tests. The reasons covered here include: (a) the usefulness of statistical significance testing in making categorical statements and testing ordinal claims; (b) researchers' dissatisfaction with the alternatives to statistical testing; and (c) the argument that statistical testing as originally conceived is a logical and sound method of statistical analysis, and persistent misuse is the fault of the researchers who misuse it rather than an indication of inherent flaws within the method. It is worth noting here that most researchers who advocate the continued use of statistical tests readily acknowledge the limitations of statistical significance testing, yet claim that for some research situations, this is one analysis of choice.

Utility in Testing Ordinal Claims

Ordinal claims are defined as those that do not specify size of effect; they specify only order or direction. Thus, "A is larger than B," and "smoking is positively correlated with lung cancer," are examples of ordinal claims because they provide directional information, but do not provide information about effect size or strength of association. Frick (1996) noted that "for quantitative claims, null hypothesis testing is not sufficient . . ., but for ordinal claims it is ideal" (p. 379). According to Abelson (1997), Frick (1996), and Greenwald, Gonzalez, Harris, and Guthrie (1996), the goal of science is not always determining size or magnitude of effect; testing ordinal claims (i.e., directional hypotheses) and making categorical statements (i.e., asserting that something important or surprising has occurred) are also important goals of science, goals for which statistical significance testing is well-suited.

Along the same lines, it has been argued that such testing can serve to maximize objectivity in the decision of whether to reject or fail to reject the null hypothesis. This means that ten independent researchers, running identical analyses on the exact same data set to test the exact same null hypothesis, will get the exact same results (i.e., p-values) and make identical decisions with regard to rejecting or failing to reject the null hypothesis (assuming identical alpha levels). Thus, proponents argue that statistical significance testing removes subjectivity from the interpretation of statistical data.

Lack of Superior Alternatives

Another argument put forward by fans of statistical tests is that proposed alternative methods, such as effect sizes, confidence intervals, and replicability analyses (all discussed below), are less informative than statistical tests, and are equally vulnerable to widespread misinterpretation (Frick, 1996; Harris, 1997). For example, Harris (1997) stated that statistical significance testing "provides useful information that is not easily gleaned from the corresponding confidence interval: degree of confidence that we have not made a Type III error [i.e., rejecting the null hypothesis in the wrong direction] and likelihood that our sample result is replicable" (p. 10). Further, as noted by Barnette and McLean (2000), results of statistical significance tests can be used to prevent researchers and research consumers from mistakenly interpreting spuriously high standardized effect sizes as meaningful. Cortina and Dunlap (1997) concluded that statistical tests and proposed alternatives such as confidence intervals each have something equally valuable to contribute to science, therefore they should be used in conjunction with each other.

Misused ≠ Misbegotten

Supporters of statistical significance tests argue that these methods are not inherently misguided or flawed; rather, years of misuse of this logical, powerful, and potentially useful tool have gradually led to its disrepute (Abelson, 1997; Cortina & Dunlap, 1997; Frick, 1996; Hagen, 1997). Hagen (1997) expressed this point eloquently:

> The logic of the [statistical test] is elegant, extraordinarily creative, and deeply embedded in our methods of statistical inference. It is unlikely that we will ever be able to divorce ourselves from that logic even if someday we decide that we want to. . . . The [statistical test] has been misinterpreted and misused for decades. This is our fault, not the fault of the [statistical test]. . . . The logic underlying statistical significance testing has not yet been successfully challenged. (p. 22)

And Abelson (1997) suggested that we

> Create a list of things that people misuse--for example, oboes, ice skates, band saws, skis, and college educations. Would you be inclined to ban them because people make errors with them? Will we want to ban effect sizes, too, when their misuse escalates? (p. 13)

Finally, Cortina and Dunlap (1997) reminded us that careful judgment is required in all areas of science, including statistical analysis, and that the "cure" for misuse and misinterpretation lies not in banning the method, but in improving our education and refining our judgment.

Arguments Against Statistical Significance Testing

Several important issues have fueled the arguments against the use of statistical significance tests. Upon reviewing the post-1994 literature, the present author found that the most often-cited and damning issues include those surrounding result replicability, sample size, what statistical significance tests actually tell us, and practical significance. Each of these interrelated issues will be covered separately below, followed by a discussion of proposed alternatives (or supplements) to the use of statistical significance tests.

The Issue of Replicability

One of the most powerful arguments against the use of statistical significance testing is that these analyses tell neither the researcher nor the research consumer anything about the replicability of a study's results. According to Thompson, the importance of replication in psychological and educational research has enjoyed increased awareness as

> Social scientists have increasingly recognized that the *single* study is inherently governed by subjective passion, that ideology frequently drives even analytic choices, and that the protection against the potentially negative consequences of these passions occurs not from feigned objectivity, but arises in the aggregate across studies from an emphasis on replication. (1994b, p. 157, emphasis in original)

The increased role of replication in educational and psychological research has been accompanied by a growing realization that statistical significance testing has severely limited utility, especially with regard to evaluating the likely replicability of study results (Cohen, 1994; Greenwald et al., 1996; Thompson, 1994b, 1995).

If the purpose of science is formulating generalizable insight based on the cumulation of findings that will generalize under stated conditions, and if the most promising strategies to fulfill this purpose emphasize interpretation based on the estimated likelihood that results will replicate, then statistical significance tests are rendered virtually useless for the underlying purpose of science. While statistical significance tests were not designed to evaluate result replicability, the real problem is that the results of these tests are often incorrectly *interpreted* as indices of replicability (Cohen, 1990). Thompson (1994b, 1995) has proposed several methods that researchers can employ to empirically assess the internal replicability of their research results; these methods include cross-validation, the bootstrap, and the jackknife.

The reason that statistical tests do *not* evaluate result replicability is that, notwithstanding common misperceptions to the contrary (Cohen, 1994), statistical tests do *not* test the probability that sample results occur in the population (Carver, 1978). In short, statistical tests assume (not test) the population, and test (not assume) the sample results! As Cohen (1994) so clearly explained, this is *not* what researchers want to do. But as he also noted, the statistical significance test "does not tell us what we want to know, and we so much want to know what we want to know that, out of desperation, we nevertheless believe that it does!" (p. 997).

The Influence of Sample Size

Another problem with statistical significance testing is that it can be circuitous, because to some extent statistical tests evaluate the size of the researcher's sample (Thompson, 1996; Zakzanis, 1998). As researchers increase their sample size, they also increase their chances of obtaining statistically significant results. Thus, as Hays argued almost 20 years ago, "virtually any study can be made to show significant results if one uses enough subjects" (1981, p. 293).

For example, suppose a group of researchers is conducting a correlational study using an alpha level of .05. With a sample size of 3, r must be as large as .997 in order to be statistically significant, while r can be as low as .276 with a sample size of 50, .196 with a sample size of 100, .088 for a sample size of 500, .062 for a sample size of 1,000, and .020 for a sample size of 10,000 (Daniel, 1998; Onwuegbuzie & Daniel, 1999). Thus, one can see how a decision to either reject or not reject the null hypothesis is largely dependent upon the researcher's sample size.

As Thompson (1998b) lamented, "Statistical testing becomes a tautological search for enough participants to achieve statistical significance. If we fail to reject, it is only because we've been too lazy to drag in enough participants" (p. 799). If any given nil null hypothesis can automatically be rejected if we just use a large enough sample, what then is the purpose of testing the hypothesis?

On the other hand, very small sample sizes can lead to failing to reject the null hypothesis when it is actually false, which is known as a Type II error. Let us consider the concept of statistical power. Cohen (1988) defined the power of a statistical test as "the probability that it will yield statistically significant results" (p. 1); that is, the probability of rejecting the null hypothesis when it is false, thereby avoiding a Type II error. Because decreasing sample size decreases the likelihood of obtaining statistical significance, power decreases as does sample size. It has been found that the average power (i.e., $1 - \beta$) of null hypothesis statistical significance tests in psychological and educational research ranges between .40 and .60; this translates to a possible 40% to 60% Type II error rate (i.e., β) (Hunter, 1997; Schmidt, 1996; Schmidt, Hunter, & Urry, 1976; Sedlmeier & Gigerenzer, 1989). Thus, due to inadequate sample sizes, many empirical inquiries in which a difference or relationship actually exists do not result in statistically significant p-values, leading the researcher to fail to reject a false null hypothesis (i.e., the statistical test results in a Type II error).

Revisiting the correlation example, this means that if the group of researchers had a sample size of 3 and obtained an r of .94 (considered to be a very large correlation by most standards), this coefficient would not be statistically significant at the .05 level due to inadequate power (Daniel, 1998; Onwuegbuzie & Daniel, 1999). Further, it has been demonstrated that psychological researchers are largely unaware of this phenomenon and tend to overestimate the power of statistical tests conducted with small samples (Schmidt et al., 1976; Tversky & Kahneman, 1971). In any event, the influence that sample size, whether large or small, has on p-values and resulting interpretation with regard to statistical significance must be recognized and further illustrates the danger of relying solely upon tests of statistical significance when analyzing data and interpreting results.

Statistical Testing Doesn't Tell Us What We Want to Know

Many researchers feel that an overemphasis on statistical significance testing detracts researchers from the primary purposes and goals of science, such as interpreting research outcomes, theory development, and formulating generalizable insight based on the cumulation of scientific findings (Kirk, 1996; Schmidt, 1996; Thompson, 1995). Thus, statistical significance testing does not usually tell us what we want to know, a point that was touched upon in the section on replicability. Indeed, Kirk (1996) reminded us that "even when a significance test is interpreted correctly, the business of science does not progress as it should" (pp. 753-754). Kirk (1996) went on:

> How far would physics have progressed if their researchers had focused on discovering ordinal relationships? What we want to know is the size of the difference between A and B and the error associated with our estimate; knowing that A is greater than B is not enough. (p. 754)

Statistical significance testing allows researchers to make a dichotomous decision with regard to their results. Either the finding is statistically significant or it isn't, and the null hypothesis is either rejected or not rejected. However, as long as researchers restrict their choices to either "statistically significant or not statistically significant," then the null hypothesis will always be false.

In psychological and educational research, it is usually of little interest to know whether the null hypothesis for a given study was rejected or not rejected (Cohen, 1990, 1994; Kirk, 1996; Thompson, 1998b). Rather, what we *really* want to know is the *magnitude* or *size* of an effect, relationship, or difference (Cohen, 1994; Kirk, 1996; Snyder & Thompson, 1998; Thompson, 1996, 1999a, 1999b); it is this knowledge that allows science to progress as the nature of effects, relationships, or differences are clarified and refined over time. While statistical significance tests may be helpful in determining the direction of relationships (i.e., testing ordinal claims) and making dichotomous decisions as regards statistical significance, they are useless with regard to producing cumulative knowledge about magnitude of relationships.

Statistical Significance vs. Practical Significance

In addition to the preceding arguments against statistical testing, many researchers are concerned with the ubiquitous practice of equating statistically significant findings with findings that are of practical significance. That is, many researchers present their data such that findings that are found to be statistically significant are also interpreted to be useful, meaningful, or important. Kirk (1996) defined the difference between statistical significance and practical significance nicely: "Statistical significance is concerned with whether a research result is due to chance or sampling variability; practical significance is concerned with whether the result is useful in the real world" (p. 746). Further, Cohen lamented:

> All psychologists know that *statistically significant* does not mean plain-English significant, but if one reads the literature, one often discovers that a finding reported in the Results section studded with asterisks implicitly becomes in the Discussion section highly significant or very highly significant, important, big! (1994, p. 1001, emphasis in original)

When researchers interpret and report their results, they have a responsibility to determine the practical importance and implications of their findings. Statistical and practical significance are two completely different concepts; a result that is not found to be statistically significant could still hold much practical significance and have meaningful implications, while a result that does happen to be statistically significant could be of little interest or importance. Result interpretation is a subjective enterprise, and researchers cannot rely upon *p*-values to inform them of the importance of their findings; *p*-values are unaware of previous research in the area, are unable to judge the implications of results for future research and practice, and are unable to evaluate a study's results in light of methodological limitations.

If not Statistical Significance Tests, Then What?

As Cohen (1994) has noted, "Don't look for a magic alternative to [statistical significance testing], some other objective mechanical ritual to replace it. It doesn't exist" (p. 1001). So what is the conscientious researcher to do? Critics of statistical significance tests have made several suggestions, with the underlying theme being for researchers to examine and interpret their data carefully and thoroughly, rather than relying solely upon *p*-values in determining which results are important enough to examine further and report in journals. Specific suggestions include the use of effect sizes, confidence intervals, and replicability analyses.

Measures of Effect Size

The reporting of effect sizes along with statistical test results in journal articles has been advocated by numerous researchers (e.g., Cohen, 1994; Kirk, 1996; Thompson, 1996, 1999b; Wilkinson & APA Task Force on Statistical Inference, 1999; Zakzanis, 1998), in addition to the APA (1994). Effect sizes are measures of the magnitude of a relationship, difference, or effect, and include variance-accounted-for effect sizes (e.g., r^2, R^2, eta^2, omega^2) and effect sizes based on standardized differences (e.g., standardized differences in means, such as Glass' Δ, Hedges' *g*, and Cohen's *d*). The variance-accounted-for effect sizes can be computed in all studies, both experimental and non-experimental, because *all* parametric analyses are part of the General Linear Model and are correlational. The effect sizes based on standardized differences can be computed when the researcher is interested in differences between group means, such as the experimental group mean and the control group mean.

As noted above, statistical significance tests allow the researcher to test directional hypotheses and make categorical statements; these tests do not, however, provide information about strength or magnitude of an association or effect, information which *is* provided by effect sizes. In short, statistical significance testing and effect size analyses differ in that the former emphasizes direction, while the latter emphasizes magnitude. Herein lies the major advantage of effect size statistics. In addition, reporting effect sizes for research outcomes enables readers to evaluate the stability of results across studies and facilitates the use of meta-analyses in future research. Further, reporting effect sizes can make research results more understandable, thereby aiding in result interpretation.

Cohen (1988) argued for the importance of researchers to interpret their calculated effect size values in light of the specific study. Just as with *p*-values, the judgment regarding the practical significance of a given effect size value depends on contextual factors such as the design of the particular study, the researcher's personal value system, societal concerns, and the research questions under study (Snyder & Lawson, 1993). Thus, an effect size of .50 may be interpreted as highly practically significant by one researcher conducting a particular study, while a different researcher (with different values and interests) conducting a different study (examining different variables with a different design) may consider this effect size as very small, and therefore as having little practical significance. For detailed information on computing and interpreting effect sizes, the reader is referred to writings by Cohen (1988), Kirk (1996), Rosenthal (1996), Rosnow and Rosenthal (1996), Snyder and Lawson (1993), and Snyder and Thompson (1998). Thompson (in press) proposes a framework for conceptualizing effect sizes and also proposes a new effect size: "corrected *d*."

Confidence Intervals

The use of confidence intervals around observed differences or computed effect sizes in research studies also has been recommended by numerous researchers (e.g., Cohen, 1990, 1994; Hunter, 1997; Kirk, 1996; Schmidt, 1996; Wilkinson & APA Task Force on Statistical Inference, 1999). Arguments for the use of confidence intervals include: (a) they are easy to compute, requiring no more information than that required for a statistical test; (b) they provide a range of values within which the true effect is likely to lie; (c) they are just as useful as statistical significance tests for deciding whether an observed difference is due to chance or sampling variability; and (d) they facilitate the interpretation of results in terms of practical and useful significance (i.e., whether the results are trivial, useful, or important). Journal editors, researchers, and other readers may find the graphic presentation of confidence intervals to be especially helpful.

Statistical computer packages (e.g., SPSS) have the capability to construct confidence intervals, thereby making their use and interpretation realistic for researchers familiar with this software and making the request for the use of these supplementary data presentation methods a reasonable one. The level of confidence associated with a given confidence interval in a given study depends on the alpha level used for that particular study. For example, if a researcher uses the .05 level of significance (α) in a given study, then the corresponding level of confidence used for constructing the confidence interval will be 95%.

Regarding the interpretation of confidence intervals, the relationship between the statistical precision of the estimate and the width of the confidence interval can be described thus: The narrower the confidence interval, the more precise the estimate (Hinkle, Wiersma, & Jurs, 1998). As sample size increases, the width of the interval will decrease if other factors are held constant, resulting in more statistical precision (i.e., the accuracy with which a confidence interval can be used to estimate a population parameter). Like the use of effect sizes, reporting confidence intervals helps researchers and readers to discern the stability of results across studies (Schmidt, 1996; Wilkinson & APA Task Force on Statistical Inference, 1999).

The APA Task Force on Statistical Inference (1999) (a) said that effect size reporting was "essential" and (b) strongly recommended the use of confidence intervals. A logical combination of these recommendations involves the reporting of confidence intervals for effect sizes themselves. However, the computation of confidence intervals for effect sizes requires the use of both sophisticated computer syntax and less familiar statistical distributions called "noncentral" distributions (e.g., noncentral *t*, noncentral *F*). Fortunately, Cumming and Finch (2001) and Smithson (2001) provide (a) understandable explanations of these applications and (b) user-friendly SPSS syntax files that implement the applications.

Replicability Analyses

Finally, the limitations of statistical tests point to the importance of either internal or external replicability analyses, which provide valuable information that statistical tests cannot (e.g., Cohen, 1994; Levin & Robinson, 1999; Robinson & Levin, 1997; Thompson, 1994b, 1995). While only external analyses invoke true replication, few researchers conduct such analyses due to the immense amount of time and effort that these analyses require. The alternative is internal replication, which can evaluate the *likely* replicability of extant study results. Internal replication methods include cross-validation, the jackknife, and the bootstrap. Although these methods are not without their limitations (Levin & Robinson, 1999; Robinson & Levin, 1997) and should not be considered substitutes for true external replication, they are certainly preferable to doing nothing at all to evaluate replicability, which is what many people do while erroneously believing that statistical tests evaluate replicability. For guidance in conducting these analyses, see Thompson (1994b, 1995).

Conclusion

The present paper has presented some extreme views on both sides of the statistical significance debate, but perhaps the most practical goal for all is compromise. If this goal were realized, statistical tests would not be completely banned, but would be routinely supplemented with accurate reports of effect size, confidence intervals, and replicability analyses. These analytic methods provide the researcher (and research consumers) with unique and essential information that statistical significance tests simply cannot, perhaps with the magnitude of effect information provided by measures of effect size being the most important. While recognizing that all data-analytic methods have their limitations and that their utility is often situation-specific, the use of effect size indices, confidence intervals, and replicability analyses in conjunction with statistical significance tests is recommended in order to facilitate comprehensive and accurate interpretation of research data.

As Shrout (1997) noted, "Significance testing has become a habit that is difficult to break" (p. 1). Maybe we don't need to completely break this habit, but we do need to practice it more responsibly, in a manner that furthers scientific knowledge. It is perfectly acceptable for researchers to use statistical tests if they feel compelled to do so and if they interpret the results of these tests accurately, but they also must conduct more informative analyses with their data. Trying to build a science *solely* on probability values and ordinal claims is a time- and energy-wasting endeavor and limits the cumulation of scientific knowledge.

It is essential to continue to evaluate the utility of statistical testing (in addition to other methods, including the alternatives discussed here) as applied to psychological and educational research. To be sure, familiarity with these considerations becomes all the more important in light of the persistent and widespread criticisms of statistical significance testing, especially as these issues slowly begin to change the editorial policies of psychological and educational journals. Perhaps the most compelling reason for researchers to be knowledgeable of these issues is the set of well-reasoned guidelines put forth by the APA Task Force on Statistical Inference (Wilkinson & APA Task Force on Statistical Inference, 1999). The Task Force's proposed guidelines will likely find their way into the next edition of the *Publication Manual*, making very immediate the need to understand how to compute and interpret alternative indices.

The requirement to report measures of effect size (and possibly additional measures) would have implications not only for researchers submitting their results for publication, but also for how graduate-level statistics courses are taught. As the APA and journal editors emphasize effect size reporting, so must educators increase the emphasis on effect sizes in the graduate statistics courses that they teach, in addition to the other supplemental analyses considered here. Not knowing how to compute or interpret effect sizes, confidence intervals, and replicability analyses may be a substantial factor contributing to the resistance to report these indices in manuscript submissions. Thus, it would seem that reform must start in the classroom, where students must be taught how to conduct these analyses competently and confidently, thereby increasing the likelihood that they will feel able to use these methods when they prepare their own work for publication.

References

Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science, 8*, 12-15.

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

Anderson, D. R., Burnham, K. P., & Thompson, W. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *Journal of Wildlife Management, 64*, 912-923.

Barnette, J. J., & McLean, J. E. (2000, April). *Use of the significance test as protection against spuriously high standardized effect sizes: Introduction to the protected effect size.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Boring, E. G. (1919). Mathematical vs. scientific importance. *Psychological Bulletin, 16*, 335-338.

Burdenski, T. (1999, January). *A review of the latest literature on whether statistical significance tests should be banned.* Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX. (ERIC Document Reproduction Service No. ED 427 084)

Carver, R. (1978). The case against statistical significance testing. *Harvard Educational Review, 48*, 378-399.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304-1312.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997-1003.

Cortina, J. M., & Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods, 2*, 161-172.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions.

Educational and Psychological Measurement, 61, 532-574.

Daniel, L. G. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. *Research in the Schools, 5*(2), 23-32.

Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods, 1,* 379-390.

Greenwald, A. G., Gonzalez, R., Harris, R. J., & Guthrie, D. (1996). Effect sizes and *p* values: What should be reported and what should be replicated? *Psychophysiology, 33,* 175-183.

Hagen, R. L. (1997). In praise of the null hypothesis statistical test. *American Psychologist, 52,* 15-24.

Harris, R. J. (1997). Significance tests have their place. *Psychological Science, 8,* 8-11.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.

Heldref Foundation. (1997). Guidelines for contributors. *Journal of Experimental Education, 65,* 95-96.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences* (4th ed.). Boston: Houghton Mifflin.

Hubbard, R., & Ryan, P. A. (2000). The historical growth of statistical significance testing in psychology--and its future prospects. *Educational and Psychological Measurement, 60,* 661-681.

Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science, 8,* 3-7.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56,* 746-759.

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement, 61,* 213-218.

Krantz, D. H. (1999). The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association, 94,* 1372-1381.

Kupfersmid, J. (1988). Improving what is published: A model in search of an editor. *American Psychologist, 43,* 635-642.

Levin, J. R., & Robinson, D. H. (1999). Further reflections on hypothesis testing and editorial policy for primary research journals. *Educational Psychology Review, 11,* 143-155.

Loftus, G. R., & Masson, M. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review, 1,* 476-490.

McLean, J. E., & Ernest, J. M. (1998). The role of statistical significance testing in educational research. *Research in the Schools, 5*(2), 15-22.

McLean, J. E., & Kaufman, A. S. (2000). Editorial: Statistical significance and other changes to *Research in the Schools. Research in the Schools, 7*(2), 1-2.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46,* 806-834.

Morrison, D. E., & Henkel, R. E. (Eds.). (1970). *The significance test controversy.* Chicago: Aldine.

Murphy, K. R. (1997). Editorial. *Journal of Applied Psychology, 82,* 3-5.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5,* 241-301.

Nix, T. W., & Barnette, J. J. (1998). The data analysis dilemma: Ban or abandon. A review of null hypothesis significance testing. *Research in the Schools, 5*(2), 3-14.

Onwuegbuzie, A. J., & Daniel, L. G. (1999, November). *Uses and misuses of the correlation coefficient.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL. (ERIC Document Reproduction Service No. ED 437 399)

Robinson, D., & Levin, J. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher, 26*(5), 21-26.

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*(4), 37-59.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276-1284.

Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1,* 331-340.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods, 1,* 115-129.

Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37-64). Mahwah, NJ: Erlbaum.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61,* 473-485.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105,* 309-316.

Shibley Hyde, J. (2001). Reporting effect sizes: The roles of editors, textbook authors, and publication manuals. *Educational and Psychological Measurement, 61*, 225-228.

Shrout, P. E. (1997). Should significance tests be banned? Introduction to a special section exploring the pros and cons. *Psychological Science, 8*, 1-2.

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement, 61*, 605-632.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61*, 334-349.

Snyder, P. A., & Thompson, B. (1998). Use of tests of statistical significance and other analytic choices in a school psychology journal: Review of practices and suggested alternatives. *School Psychology Quarterly, 13*, 335-348.

Svyantek, D. J., & Ekeberg, S. E. (1995). The earth is round (so we can probably get there from here). *American Psychologist, 50*, 1101.

Thompson, B. (Ed.). (1993). Theme issue: Statistical significance testing in contemporary practice [Special issue]. *Journal of Experimental Education, 61*(4).

Thompson, B. (1994a). Guidelines for authors. *Educational and Psychological Measurement, 54*, 837-847.

Thompson, B. (1994b). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality, 62*, 157-176.

Thompson, B. (1995). Exploring the replicability of a study's results: Bootstrap statistics for the multivariate case. *Educational and Psychological Measurement, 55*, 84-94.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*(2), 26-30.

Thompson, B. (1998a, April). *Five methodology errors in educational research: The pantheon of statistical significance and other faux pas.* Invited address presented at the annual meeting of the American Educational Research Association, San Diego. (ERIC Document Reproduction Service No. ED 419 023)

Thompson, B. (1998b). In praise of brilliance: Where that praise really belongs. *American Psychologist, 53*, 799-800.

Thompson, B. (1998c). Review of *What if there were no significance tests?* by L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.). *Educational and Psychological Measurement, 58*, 332-344.

Thompson, B. (1999a). If statistical significance tests are broken/misused, what practices should supplement or replace them? *Theory and Psychology, 9*, 165-181.

Thompson, B. (1999b). Improving research clarity and usefulness with effect size indices as supplements to statistical significance tests. *Exceptional Children, 65*, 329-337.

Thompson, B. (1999c). Statistical significance tests, effect size reporting, and the vain pursuit of pseudo-objectivity. *Theory and Psychology, 9*, 193-199.

Thompson, B. (in press). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling and Development.*

Tryon, W. W. (1998). The inscrutable null hypothesis. *American Psychologist, 53*, 796.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Vacha-Haase, T. (2001). Statistical significance should not be considered one of life's guarantees: Effect sizes are needed. *Educational and Psychological Measurement, 61*, 219-224.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Zakzanis, K. K. (1998). Brain is related to behavior ($p < .05$). *Journal of Clinical and Experimental Neuropsychology, 20*, 419-427.

JOURNAL SUBSCRIPTION FORM

This form can be used to subscribe to *RESEARCH IN THE SCHOOLS* without becoming a member of the Mid-South Educational Research Association. It can be used by individuals and institutions.

◆

Please enter a subscription to *RESEARCH IN THE SCHOOLS* for:

Name: ____________________

Institution: ____________________

Address: ____________________

____

			COST
Individual Subscription ($30 per year)	Number of years	____	____
Institutional Subscription ($40 per year)	Number of years	____	____
Foreign Surcharge ($30 per year, applies to both individual and institutional subscriptions)	Number of years	____	____
Back issues for Volumes 1 - 7 ($30 per Volume)	Number of Volumes	____	____
TOTAL COST:			____

MAKE CHECKS PAYABLE TO MSERA
SEND FORM AND CHECK TO:

Dr. James E. McLean, Co-Editor
RESEARCH IN THE SCHOOLS
Warf-Pickle Hall, Room 418
Eastern Tennessee State University
Box 70685
Johnson City, TN 37614-1709

MID-SOUTH EDUCATIONAL RESEARCH ASSOCIATION

The Mid-South Educational Research Association (MSERA) was founded in order to encourage quality educational research in the Mid-South and to promote the application of quality educational research in schools. Members of MSERA share interests in educational research, development, and evaluation. While most members are from institutions of higher education, many others represent state departments of education, public and private agencies, and public school systems. Graduate students comprise a significant portion of the membership. A majority of MSERA members are from the six states represented by the organization, but others are from many other states and several foreign countries. The MSERA is the largest regional educational research organization in the country.

The organization provides several services for its members. The annual meeting, held every November, offers many formal and informal opportunities for professional development through special training courses, sharing of research findings, and programmatic interests with colleagues. Members receive a subscription to *RESEARCH IN THE SCHOOLS* and the *Mid-South Educational Researcher*. The MSERA also provides recognition and cash rewards for its outstanding paper, an outstanding dissertation, and professional service.

MSERA Membership/Renewal Form

(Please print or type)

Name ______________________________

Organization ______________________________

Address ______________________________

____

Telephone Work: ____________________

Home: ____________________

Fax: ____________________

e-mail: ______________________________

Amount Enclosed: MSERA 2002 Membership ($25 professional, $15 student) $________

MSER Foundation Contribution $________

TOTAL $________

Make check out to MSERA and mail to:

Dr. Ernest Rakow
University of Memphis
Ball Education Bldg., Room 100
Memphis, TN 38152-6010

Research in the Schools
Mid-South Educational Research Association
Warf-Pickle Hall, Room 418
Eastern Tennessee State University
Johnson City, TN 37614-1709

PRSRT STD.
U. S. POSTAGE
PAID
PERMIT # 3029
BIRMINGHAM, AL

TM034711

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

NOTICE

Reproduction Basis

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (3/2000)